

Spectral estimation for diffusions with random sampling times

Jakub Chorowski and Mathias Trabs*

Humboldt-Universität zu Berlin and Université Paris-Dauphine

Abstract

The nonparametric estimation of the volatility and the drift coefficient of a scalar diffusion is studied when the process is observed at random time points. The constructed estimator generalizes the spectral method by Gobet, Hoffmann and Reiß [Ann. Statist. 32 (2006), 2223-2253]. The estimation procedure is optimal in the minimax sense and adaptive with respect to the sampling time distribution and the regularity of the coefficients. The proofs are based on the eigenvalue problem for the generalized transition operator. The finite sample performance is illustrated in a numerical example.

MSC2010 subject classification: Primary 62M05; Secondary 60J60, 62G99, 62M15.

Key words and phrases: Ergodic diffusion processes, generalized transition operator, Lepski's method, minimax optimal convergence rates, nonparametric estimation, random sampling.

1 Introduction

For decades diffusion models are used to describe the dynamics of continuous stochastic processes, for instance, stock prices in econometrics or particle movements in biology and physics. The statistical properties of diffusion models depend essentially on the observation scheme, where it is natural to assume discrete observations of the process. Mostly, equidistant observations are studied in the literature, distinguishing between high-frequent and low-frequent observations, depending whether the observation distance tends to zero or remains fixed. A summary of parametric methods is given by Aït-Sahalia [2]. Nonparametric estimation methods are surveyed by Fan [13].

As argued by Aït-Sahalia and Mykland [3], assuming equidistant observations might however not be realistic in many applications and random sampling times should be instead considered. For parametric estimation problems Aït-Sahalia and Mykland [3, 4] have shown that random sampling has a strong effect on the statistical problem and the performance of estimators. Naturally, the question arises how nonparametric estimators can be constructed for random sampling times and whether their (asymptotic) behavior is similar or worse than for equidistant observations.

In order to study the nonparametric estimation of the drift and the volatility coefficient of the diffusion when the process is observed at random times, we generalize the low-frequency results by Gobet et al. [14]. As they do, we consider a reflected scalar diffusion on a one-dimensional interval. On the one hand, this allows to avoid technical difficulties and to present more transparent proofs when investigating spectral properties of the transition semigroup. On the other hand, diffusions with reflecting barriers have rich applications. In the finance and economics literature reflected

*The authors thank Markus Reiß for helpful comments and discussions. J.C. was financially supported by the Deutsche Forschungsgemeinschaft (DFG) RTG 1845 "Stochastic Analysis with Applications in Biology, Finance and Physics". M.T. acknowledges financial support by the DFG through the CRC 649 "Economic Risk" and the research fellowship TR 1349/1-1. The main part of the paper was carried out while M.T. was employed at the Humboldt-Universität zu Berlin.

diffusions are used for currency exchange rate target-zone models, in which the exchange rate is allowed to float within two barriers enforced by the monetary authority c.f. [6, 19, 31]. Reflected diffusions also appear as the payoff of the so-called “Russian Options”, c.f. Shepp and Shiryaev [27]. Among applications in mathematical biology, we recall models for population dynamics in which the total number of individuals is affected by oppositely acting forces, e.g., spontaneous growth and immigration on the one hand and random harvesting or predation on the other, c.f. [25]. Finally reflected Brownian motion have been shown to describe queueing models experiencing heavy traffic, see [16, 18]. In all these models the observation times might not be equidistantly distributed. For instance, they depend on trading times for finance applications or measurement times of the biologist.

By the compactness of the interval and the reflecting boundary, the diffusion is ergodic and admits a spectral gap. Our procedure relies on a representation of the coefficients in terms of the invariant measure and the first non-trivial eigenpair of the infinitesimal generator of the diffusion. This spectral identification method was introduced in Hansen et al. [15] and has been further studied by [10]. It is crucial that the eigenpair is determined by the transition operator of the time changed diffusion, where the time change is given by the sampling distribution and the Laplace transform of the sampling distribution. The former can be estimated by a wavelet projection method and latter by classical empirical process theory. As a side product of our analysis we clarify some aspects of the estimator and the proofs by Gobet et al. [14]. In particular, in order to stabilize the estimator against large stochastic errors a truncation with an in practice unknown threshold value was needed, which we could omit.

Moreover, we show that Lepski’s method can be applied to chose the projection level in a data-driven way. This allows to adapt on the unknown Sobolev regularity of the drift and volatility coefficients of the diffusion. The first adaptive estimator based on low-frequency observations of a diffusion process has been constructed only recently in Söhl and Trabs [28]. Considering diffusion on the whole real line, this first result is restricted to a diffusion with constant volatility, simplifying the whole estimation problem, we do not need any additional restrictions on the drift or the volatility.

We prove that the estimators achieve minimax optimal convergence rates. The adaptive estimator only loses a logarithmic factor. In view of the *cost of randomness* determined by Aït-Sahalia and Mykland [4], it might be surprising that the convergence rates do not depend on the sampling distribution and coincide in fact with the nonparametric rates of the low-frequency setting. In that sense, our method is also adaptive with respect to the unknown sampling distribution. As one can see clearly from simulations, there is, however, a large *cost of ignoring the randomness* in the misspecified case where one applies the low-frequency estimator to randomly sampled observations using the average time step as observations distance.

The paper is organized as follows: In Section 2 we introduce the diffusion with reflected boundaries, our basic assumptions and the main properties of the process. The estimators are constructed in Section 3. The main results on the convergence rates are stated and discussed in Section 4. The adaptive estimator is constructed in Section 5. The finite sample performance of the method is illustrated in a small simulation study in Section 6. The proofs of the upper and lower bounds as well as for the Lepski method are postponed to Sections 7, 8 and 9, respectively. Finally, some results on the stability of the eigenvalue problems are presented in the appendix.

2 The model

Without loss of generality we can consider the unit interval $[0, 1]$ for the reflecting diffusion. For a measurable and bounded drift function $b: [0, 1] \rightarrow \mathbb{R}$ and a continuous volatility function $\sigma: [0, 1] \rightarrow \mathbb{R}_+$ let the process $X = \{X_t : t \geq 0\}$ be given by the stochastic differential equation

$$\begin{aligned} dX_t &= b(X_t) dt + \sigma(X_t) dW_t + v(X_t) dY_t(X), \\ X_0 &= x_0, \text{ and for all } t \geq 0 \ X_t \in [0, 1], \end{aligned} \tag{1}$$

where x_0 is a random variable on $[0, 1]$, $W = \{W_t : t \geq 0\}$ is a standard Brownian motion, $v: [0, 1] \rightarrow \mathbb{R}$ satisfies $v(0) = 1, v(1) = -1$, and Y , which is part of the solution, is a non-anticipative continuous non-decreasing process increasing only when $X_t \in \{0, 1\}$. By the Engelbert-Schmidt theorem boundedness of the drift coefficient together with the volatility function being continuous and strictly positive ensure that (1) has a weak solution, see Rozkosz and Słomiński [26, Thm. 4.1]. We denote by $\mathbb{P}_{\sigma, b}$ the law of this solution on the canonical space $\Omega = C(\mathbb{R}_+, [0, 1])$ of continuous functions equipped with the topology of uniform convergence on compact subsets and endowed with its Borel σ -field \mathcal{F} .

For $N \in \mathbb{N}$ our observations are given by

$$(0, X_0), (\tau_1, X_{\tau_1}), \dots, (\tau_N, X_{\tau_N}) \in [0, \infty) \times [0, 1]$$

where τ_1, \dots, τ_N is an increasing sequence of random time points. For convenience we write $\tau_0 = 0$.

Assumption 1. *Let the observation distances*

$$\Delta_n := \tau_n - \tau_{n-1}, \quad n = 1, \dots, N,$$

be an independent and identically distributed sequence of strictly positive random variables with law

$$\gamma \in \Gamma := \Gamma(I, \alpha) := \{\gamma \text{ probability measure on } \mathbb{R}_+ : \gamma(I) \geq \alpha\}$$

for some compact interval $I \subset (0, \infty)$ and some $\alpha \in (0, 1]$. Let Δ_n be independent of the diffusion process X .

This condition on the sampling distributions is very weak. For every given positive distribution γ there are I, α such that $\gamma \in \Gamma(I, \alpha)$. The only restrictions are that the set Γ has to be bounded in the right sense, since we will derive uniform rates in this class, and we have to exclude distributions that concentrate at zero. The latter condition is natural because otherwise the observations would be of high-frequency type which would require a completely different analysis.

Example 2.

- (i) The special case of the low-frequency observations is covered by setting $\tau_n = n\Delta$ for some fixed deterministic $\Delta > 0$. Then the sampling distribution is given by the Dirac measure in Δ , that is $\Gamma = \{\delta_\Delta\}$.
- (ii) If the observation times are governed by a Poisson process, the waiting time to the next observation is exponentially distributed, that is $\gamma = \text{Exp}(\lambda)$ for some intensity $\lambda > 0$. In this case we can choose $\Gamma = \{\text{Exp}(\lambda) : \lambda \in \Lambda\}$ for any bounded set $\Lambda \subset (0, \infty)$.

To state the assumptions on the diffusion coefficients, we denote the $L^2([0, 1])$ Sobolev space of order $s > 0$ by $H^s := H^s([0, 1])$. Furthermore, let $H_b^s \subset H^s$ be the subset of bounded functions with Sobolev regularity s . Note that $H_b^s = H^s$ for $s > 1/2$ by the Sobolev embeddings.

Assumption 3. *For $s > 1$ and constants $d, D > 0$ let $(\sigma, b) \in \Theta_s$ where*

$$\Theta_s := \Theta_s(d, D) = \left\{ (\sigma, b) \in H^s \times H_b^{s-1} : \|\sigma^2\|_{H^s} \leq D, \|b\|_{H^{s-1}} \leq D, \inf_x \sigma(x) \geq d \right\}.$$

In particular, $(\sigma, b) \in \Theta_s$ ensures the existence of a weak solution of (1). As shown by Gobet et al. [14] the compactness of $[0, 1]$ and the reflecting boundary conditions imply that X has a spectral gap and thus it is geometrically ergodic and admits an invariant measure μ . Focusing on asymptotic results, we can suppose that the initial value x_0 is distributed according to μ . Assumption 3 implies that μ has the Lebesgue density, abusing notation denoted by μ as well,

$$\mu(x) := \mu_{\sigma, b}(x) = C_0 \sigma^{-2}(x) \exp \left(\int_0^x 2b(y) \sigma^{-2}(y) dy \right), \quad x \in [0, 1], \quad (2)$$

for some normalizing constant $C_0 > 0$, cf. Bass [8, Chap. 4] or Karlin and Taylor [17, Chap. 15, Sect. 6]. It is easy to see that the regularity assumptions on b and σ imply that $\mu \in H^s$, which will be essential for the analysis of the estimators. From the explicit formula for μ moreover follows that there are constants $0 < c < C$ such that $c \leq \mu_{\sigma,b} \leq C$ for any $(\sigma, b) \in \Theta_s$. Consequently, $L^2(\mu)$ with the inner product

$$\langle f, g \rangle_\mu := \int_0^1 f(x)g(x)\mu(x)dx$$

is a Hilbert space equivalent to $L^2([0, 1])$.

Noting that reflection corresponds to Neumann boundary conditions, the infinitesimal generator $L = L_{\sigma,b}$ of the diffusion X is an unbounded, densely defined operator on $L^2([0, 1])$ satisfying

$$\begin{aligned} Lf(x) &= b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x), \\ \text{dom}(L) &= \{f \in H^2([0, 1]) : f'(0) = f'(1) = 0\}. \end{aligned}$$

Furthermore, seen as an operator on the Hilbert space $L^2(\mu)$, the generator L is an elliptic, self-adjoint operator with compact resolvent, see Chatelin [9, Example 4.21]. Consequently it has a pure point spectrum $\sigma(L) = \{v_k : k = 0, 1, \dots\}$ and the corresponding eigenfunctions u_k form an $L^2(\mu)$ orthogonal basis. Its largest eigenvalue v_0 equals 0 with constant corresponding eigenfunction. All other eigenvalues are negative and we assume that they are ordered with respect to their multiplicities $0 > v_1 \geq v_2 \geq \dots$. As shown in [14, Lemma 6.1], the eigenvalue v_1 is simple and the eigenfunction u_1 can be chosen strictly increasing.

3 Estimation method

3.1 Spectral identification

The main idea used for the construction of the spectral estimators in [14] is that the coefficients of a stationary diffusion process can be expressed in terms of the invariant density μ and any nontrivial eigenpair (v_k, u_k) , $k \geq 1$. Indeed, expressing the invariant measure in terms of the speed measure together with the Neumann boundary conditions yields, cf. [14, Sect. 3.1],

$$\sigma^2(x) = \frac{2v_k \int_0^x u_k(y)\mu(y)dy}{u'_k(x)\mu(x)}, \quad (3)$$

$$\begin{aligned} b(x) &= \frac{v_k u_k(x)}{u'_k(x)} - \frac{\sigma^2(x)u''_k(x)}{2u'_k(x)} \\ &= v_k \frac{u_k(x)u'_k(x)\mu(x) - u''_k(x) \int_0^x u_k(y)\mu(y)dy}{u'_k(x)^2 \mu(x)}. \end{aligned} \quad (4)$$

Applying the ergodicity, it is easy to estimate the invariant measure μ . To recover an eigenpair of the generator, Gobet et al. [14] have used discrete equidistant observations, i.e. $\Delta_n = \Delta$ for some fixed $\Delta > 0$, to construct a matrix estimator of the transition operator $P_\Delta = e^{\Delta L}$. Noting that P_Δ shares eigenfunctions with the generator L while its eigenvalues are $e^{\Delta v_k}$, $k = 0, 1, \dots$, they have obtained estimators of (v_k, u_k) . We will generalize these results taking into account the random observation times τ_1, \dots, τ_N .

Similar to the transition operator P_Δ we introduce the *generalized transition operator* R on $L^2(\mu)$ given by

$$Rf(x) = \mathbb{E}_{\sigma,b,\gamma} [f(X_\tau) | X_0 = x], \quad x \in [0, 1], \quad (5)$$

where τ is a random variable with distribution γ being independent of the process X . The crucial insight is that for any eigenpair (v_k, u_k) of the generator we have

$$Ru_k(x) = \mathbb{E}_{\sigma, b, \gamma} [\mathbb{E}_{\sigma, b, \gamma} [P_t u_k | \tau = t]] = \mathbb{E}_{\gamma} [e^{\tau v_k}] u_k(x) = \underbrace{\mathcal{L}_{\gamma}(-v_k)}_{=:\kappa_k} \cdot u_k(x), \quad (6)$$

where

$$\mathcal{L}_{\gamma}(z) := \int_0^{\infty} e^{-tz} \gamma(dt), \quad z \in \mathbb{R}_+, \quad (7)$$

is the Laplace transform of γ . Consequently, R is a compact operator with eigenvalues $1 = \kappa_0 > \kappa_1 > \kappa_2 \geq \kappa_3 \geq \dots > 0$. In the functional calculus sense we obtain

$$R = \mathcal{L}_{\gamma}(-L).$$

Therefore, we can estimate the eigenpairs (v_k, u_k) using the spectral properties of R . Since the sampling distribution γ is unknown, we need to estimate the Laplace transform from the observations $(\Delta_n)_{n=1, \dots, N}$.

Example 2 (continued). (i) For $\Delta_n \equiv \Delta$ for some fixed $\Delta > 0$ we have $Rf = P_{\Delta}f$ and $\mathcal{L}_{\gamma}(z) = e^{-\Delta z}$, $z \geq 0$. We thus exactly recover the situation studied in [14].

(ii) If $\Delta_n \sim \text{Exp}(\lambda)$, then the Laplace transform is given by $\mathcal{L}_{\gamma}(z) = \int_0^{\infty} \lambda e^{-t(z+\lambda)} dt = \frac{\lambda}{z+\lambda}$, $z \geq 0$ and the operator R is the resolvent of the generator L .

The distribution of the eigenvalues of the operator R is inherited from the generator L and the sampling distribution γ . More precisely, we obtain the following lemma whose proof is postponed to Section 7.1.

Lemma 4. Grant Assumptions 1 and 3. The spectral gap, that is $\inf_{i \neq 1} |\kappa_i - \kappa_1|$, and the eigenvalues of the generalized transition operator R have a lower bound uniform in $(\sigma, b) \in \Theta_s$ and $\gamma \in \Gamma$.

3.2 Construction of the estimators

Let us fix some notation. We will write $f \lesssim g$ (resp. $g \gtrsim f$) when $f \leq C \cdot g$ for some universal constant $C > 0$. $f \sim g$ is equivalent to $f \lesssim g$ and $g \lesssim f$. Let (ψ_{λ}) , with multi-indices $\lambda = (j, k)$, be an L^2 -orthonormal regular wavelet basis of $L^2([0, 1])$. The corresponding approximation spaces are given by

$$V_J := \overline{\text{span}}\{\psi_{\lambda} : |\lambda| = |(j, k)| := j \leq J\}.$$

The L^2 -orthogonal and the $L^2(\mu)$ -orthogonal projections onto V_J are denoted by π_J and π_J^{μ} , respectively.

In fact, the approximation spaces do not necessarily need to be generated by wavelets. We only require that V_J , $J \in \mathbb{N}$, satisfy Jackson and Bernstein type inequalities with respect to the Sobolev spaces H^s , that is for all $0 \leq t \leq s$, $f \in H^s$ and $g \in V_J$

$$\|(I - \pi_J)f\|_{H^t} \lesssim 2^{-J(s-t)} \|f\|_{H^s} \quad \text{and} \quad \|g\|_{H^j} \lesssim 2^{Jj} \|g\|_{L^2}, \quad j = 1, 2, \quad (8)$$

and additionally we need the uniform bound

$$\left\| \sum_{|\lambda| \leq J} \psi_{\lambda}^2 \right\|_{\infty} \lesssim \dim(V_J) = 2^J. \quad (9)$$

It follows from the well known properties of wavelets that (8) and (9) are satisfied.

Remark 5. Since the eigenfunctions of the generator of the reflected Brownian motion are given by the trigonometric functions, it seems to be attractive to choose V_J as the closure of the span of the first 2^J orthogonal trigonometric basis functions, which however does not fulfill (8). If the drift and the volatility function satisfy the stronger Hölder regularity assumption $\|\sigma^2\|_{C^s}, \|b\|_{C^{s-1}} \leq D$, where $\|\cdot\|_{C^s}$ denotes the Hölder norm, then we can obtain the same bounds on the mean L^2 estimation error under a weaker version of Jackson's inequality, namely

$$\|(I - \pi_J)f\|_{L^2} \lesssim 2^{-Js} \|f\|_{C^s}.$$

This inequality is satisfied for the trigonometric basis. Furthermore Bernstein's inequality can be easily checked and (9) is trivially fulfilled. The same applies to the B-spline basis, that satisfies above conditions with the weakened Jackson inequality (see [11] and [12]).

After having fixed the basis functions and the corresponding approximation spaces V_J , there is a one-to-one correspondence between a linear operator $A: V_J \rightarrow V_J$ on the finite dimensional space V_J and its matrix representation $(A_{\lambda, \lambda'}) \in \mathbb{R}^{\dim V_J \times \dim V_J}$ with $A_{\lambda, \lambda'} := \langle \psi_\lambda, A\psi_{\lambda'} \rangle$. To simplify the notation, we will throughout use A to denote the operator as well as its representation matrix.

Using the ergodicity of the diffusion X and the independence of X and $(\Delta_n)_n$, the sequence $(X_{\tau_n})_n$ is ergodic, too. The natural estimator for the invariant measure is therefore the empirical measure

$$\mu_N = \frac{1}{N+1} \sum_{n=0}^N \delta_{X_{\tau_n}}.$$

To regularize μ_N , we define the projection estimator

$$\hat{\mu}_J(x) := \sum_{|\lambda| \leq J} \langle \psi_\lambda, \mu_N \rangle \psi_\lambda(x) \quad \text{with} \quad \langle \psi_\lambda, \mu_N \rangle := \frac{1}{N+1} \sum_{n=0}^N \psi_\lambda(X_{\tau_n})$$

for a projection level $J \in \mathbb{N}$. We proceed similarly to Gobet et al. [14]. Extending the matrix estimator of the transition semigroup, we introduce the matrix estimator $\hat{R}_J = (\hat{R}_{\lambda, \lambda'})$ of the action of the operator R from (5) on the wavelet basis with respect to the scalar product $\langle \cdot, \cdot \rangle_\mu$:

$$\hat{R}_{\lambda, \lambda'} := \frac{1}{2N} \sum_{n=0}^{N-1} \left(\psi_\lambda(X_{\tau_{n+1}}) \psi_{\lambda'}(X_{\tau_n}) + \psi_{\lambda'}(X_{\tau_{n+1}}) \psi_\lambda(X_{\tau_n}) \right).$$

Since the observation times are independent from the diffusion, conditioning on τ_n , we can verify that \hat{R}_J is an unbiased estimator of the action of the operator R on the basis, that is

$$\mathbb{E}_{\sigma, b, \gamma} [\hat{R}_{\lambda, \lambda'}] = \langle \psi_\lambda, R\psi_{\lambda'} \rangle_\mu.$$

The Gram matrix $G_J = (\langle \psi_\lambda, \psi_{\lambda'} \rangle_\mu)_{\lambda, \lambda'} \in \mathbb{R}^{\dim V_J \times \dim V_J}$ is determined by $\langle v, G_J v \rangle = \langle v, v \rangle_\mu$ for all $v \in V_J \setminus \{0\}$. Hence, G_J is a restriction of the scalar product $\langle \cdot, \cdot \rangle_\mu$ to finite dimensional space V_J . It can be estimated by $\hat{G}_J = (\hat{G}_{\lambda, \lambda'})$ with

$$\hat{G}_{\lambda, \lambda'} = \frac{1}{N} \left(\frac{1}{2} \psi_\lambda(X_0) \psi_{\lambda'}(X_0) + \sum_{n=1}^{N-1} \psi_\lambda(X_{\tau_n}) \psi_{\lambda'}(X_{\tau_n}) + \frac{1}{2} \psi_\lambda(X_{\tau_N}) \psi_{\lambda'}(X_{\tau_N}) \right),$$

satisfying

$$\mathbb{E}_{\sigma, b, \gamma} [\hat{G}_{\lambda, \lambda'}] = \langle \psi_\lambda, \psi_{\lambda'} \rangle_\mu = \langle \psi_\lambda, G_J \psi_{\lambda'} \rangle.$$

Owing to $\langle v, G_J v \rangle = \langle v, v \rangle_\mu > 0$ for any $v \in V_J \setminus \{0\}$, the matrix G_J is invertible. By construction $\langle v, \hat{G}_J v \rangle$ is always non-negative and it will be strictly positive whenever the sample is sufficiently dispersed over all the interval $[0, 1]$. By ergodicity we can expect this to be a high probability event. With a Neumann series argument we can moreover bound the norm of \hat{G}_J^{-1} as stated by the following lemma, which is proven in Section 7.4.

Lemma 6. *Grant Assumption 1 and 3. On the event $\mathcal{T}_1 = \left\{ \|G_J - \hat{G}_J\|_{L^2} \leq \frac{1}{2} \|G_J^{-1}\|_{L^2}^{-1} \right\}$ the estimator \hat{G}_J is invertible and satisfies $\|\hat{G}_J^{-1}\|_{L^2} \leq 2\|G_J^{-1}\|_{L^2}$. Moreover, $\mathbb{P}_{\sigma,b,\gamma}(\Omega \setminus \mathcal{T}_1) \leq N^{-1}2^{2J}$ holds uniformly over Θ_s and Γ .*

Whenever \hat{G}_J^{-1} exists, we can consider $\hat{G}_J^{-1}\hat{R}_J$. Since \hat{R}_J is symmetric it immediately follows that $\hat{G}_J^{-1}\hat{R}_J$ is symmetric with respect to the \hat{G}_J -scalar product. Furthermore, by the Cauchy-Schwarz inequality and the inequality between geometric and arithmetic means we obtain for all $v \in V_J \setminus \{0\}$

$$\begin{aligned} \langle \hat{R}_J v, v \rangle &= \frac{1}{N} \sum_{n=0}^{N-1} v(X_{\tau_n}) v(X_{\tau_{n+1}}) \\ &\leq \frac{1}{N} \left(\sum_{n=0}^{N-1} v^2(X_{\tau_n}) \right)^{1/2} \left(\sum_{n=1}^N v^2(X_{\tau_n}) \right)^{1/2} \\ &\leq \frac{1}{N} \left(\frac{1}{2} v^2(X_0) + \frac{1}{2} v^2(X_{\tau_N}) + \sum_{n=1}^{N-1} v^2(X_{\tau_n}) \right) = \langle \hat{G}_J v, v \rangle. \end{aligned}$$

Consequently, all eigenvalues of the matrix $\hat{G}_J^{-1}\hat{R}_J$ are real and smaller than one. It is easy to check that 1 is an eigenvalue corresponding to the constant function. We define the estimator $(\hat{\kappa}_{J,1}, \hat{u}_{J,1})$ of the eigenpair (κ_1, u_1) as the eigenpair of the matrix $\hat{G}_J^{-1}\hat{R}_J$ corresponding to the biggest eigenvalue smaller than one. On the exceptional event that \hat{G}_J is not invertible, we set $\hat{\kappa}_{J,1} = 0$ and $\hat{u}_{J,1} = 1$. Furthermore we choose the estimated eigenfunction $\hat{u}_{J,1}$ normalized in L^2 .

Using $\hat{\kappa}_{J,1}$ and the identification equation $\kappa_1 = \mathcal{L}_\gamma(-v_1)$, we can estimate v_1 . The canonical estimator for the Laplace transform of γ is the Laplace transform of the empirical measure of the sampling distances $\Delta_n = \tau_n - \tau_{n-1}, n = 1, \dots, N$. Hence, we define

$$\hat{\mathcal{L}}(y) := \frac{1}{N} \sum_{n=1}^N e^{-y\Delta_n}, \quad y \in \mathbb{R}_+.$$

Due to the i.i.d. structure of (Δ_n) , the classical empirical process theory shows that $\hat{\mathcal{L}}$ estimates \mathcal{L}_γ uniformly in a neighborhood of v_1 with the parametric rate $N^{-1/2}$. Moreover, $\hat{\mathcal{L}}$ is strictly decreasing and continuous, thus invertible. We define

$$\hat{v}_{J,1} := -\hat{\mathcal{L}}^{-1}(\hat{\kappa}_{J,1}) \mathbf{1}_{\{\hat{\kappa}_{J,1} > 0\}}. \quad (10)$$

With the above definitions and in view of the identification formulas (3) and (4) we can define the plug-in estimators of the diffusion coefficients. In order to ensure integrability of our estimators, we need to stabilize against large stochastic errors. Using the prior knowledge that $(\sigma, b) \in \Theta_s$, especially $\|\sigma^2\|_\infty \leq D$ and $\|b\|_{L^2} \leq D$ for some $D > 0$, we thus define

$$\hat{\sigma}_J^2(x) = 2\hat{v}_{J,1} \frac{\int_0^x \hat{u}_{J,1}(y) \hat{\mu}_J(y) dy}{\hat{u}'_{J,1}(x) \hat{\mu}_J(x)} \wedge D, \quad (11)$$

$$\hat{b}_J(x) = \tilde{b}_J(x) \mathbf{1}_{\{\|\tilde{b}_J\|_{L^2} \leq 2D\}} \quad \text{for} \quad \tilde{b}_J(x) := \frac{\hat{v}_{J,1} \hat{u}_{J,1}(x)}{\hat{u}'_{J,1}(x)} - \frac{\hat{\sigma}_J^2(x) \hat{u}_{J,1}''(x)}{2\hat{u}'_{J,1}(x)}. \quad (12)$$

4 Minimax convergence rates

Let us now state our first main results, generalizing Theorems 2.4 and 2.5 in [14], respectively. Note that since $u'_1(0) = u'_1(1) = 0$ the function

$$[0, 1] \ni x \mapsto \frac{2v_1 \int_0^x u_1(y) \mu(y) dy}{u'_1(x) \mu(x)} = \sigma^2(x)$$

is defined in $\{0, 1\}$ via continuous extension such that the proposed estimators $\hat{\sigma}_J^2$ and \hat{b}_J might be unstable at the boundary. We restrict the L^2 -loss to an interval $[a, b] \subset [0, 1]$ for $0 < a < b < 1$ and refer to [14, Section 3.3.8] for a discussion of the boundary problem.

Theorem 7. *Grant Assumptions 1 and 3 for some $s > 1$. Let $0 < a < b < 1$. Choosing $2^J \sim N^{1/(2s+3)}$, we have*

$$\begin{aligned} \sup_{(\sigma^2, b, \gamma) \in \Theta_s \times \Gamma} \mathbb{E}_{\sigma, b, \gamma} [\|\hat{\sigma}_J^2 - \sigma^2\|_{L^2([a, b])}^2] &\lesssim N^{-2s/(2s+3)}, \\ \sup_{(\sigma^2, b, \gamma) \in \Theta_s \times \Gamma} \mathbb{E}_{\sigma, b, \gamma} [\|\hat{b}_J - b\|_{L^2([a, b])}^2] &\lesssim N^{-2(s-1)/(2s+3)}. \end{aligned}$$

The risk of $\hat{\sigma}^2$ and \hat{b} decomposes into the errors for estimating the invariant density μ and the eigenpair and (v_1, u_1) of the infinitesimal generator L of the diffusion. In view of formula (2) the invariant density inherits Sobolev regularity of degree s from the diffusion coefficients. Together with the ergodicity and the spectral gap μ can be estimated with the rate $\mathbb{E}_{\sigma, b, \gamma} [\|\hat{\mu}_J - \mu\|_{L^2}] \lesssim N^{-\frac{s}{2s+1}}$ if we choose $2^J \sim N^{-1/(2s+1)}$, cf. Proposition 11. Due to $\mathcal{L}_\gamma(-v_1) = \kappa_1$ estimating v_1 reduces to estimate the eigenvalue κ_1 of the operator R and the inverse of the Laplace transform \mathcal{L}_γ in a neighborhood of κ_1 . The latter estimation problem can be solved with standard empirical process results yielding the parametric rate $N^{-1/2}$ for $\hat{\mathcal{L}}$, see Lemma 18.

The analysis of the estimation error of the eigenpair (κ_1, u_1) of the generalized transition operator R is the most challenging ingredient of our proofs. We first restrict the eigenvalue problem to the finite dimensional space V_J , that is we find $(\kappa_{J,1}, u_{J,1}) \in \mathbb{R}_+ \times V_J$ such that

$$\langle v, Ru_{J,1} \rangle_\mu = \kappa_{J,1} \langle v, u_{J,1} \rangle_\mu \quad \text{for all } v \in V_J. \quad (13)$$

As shown in Theorem 25 the resulting approximation error $\|u_1 - u_{J,1}\|_{L^2(\mu)} + |\kappa_1 - \kappa_{J,1}|$ is controlled by the spectral gap of the operator R and the smoothness of the eigenfunction (of degree $s+1$) achieving the order of magnitude $2^{-J(s+1)}$. In the second step we approximate the finite dimensional problem (13) by a generalized symmetric eigenvalue problem for the random matrices \hat{R}_J and \hat{G}_J . We use classical a posteriori error bounds to show that the approximation error is controlled by the norm of the so called residual vector $r = (\hat{R}_J - \kappa_{J,1}\hat{G}_J)u_{J,1}$, cf. Theorem 26. $\|r\|_{L^2}$ can be bounded by the matrix approximation errors $\|(\hat{R}_J - R_J)u_{J,1}\|_{L^2}$ and $\|(\hat{G}_J - G_J)u_{J,1}\|_{L^2}$ that tend to zero by the mixing property of the Markov chain $(X_{\tau_n})_n$. A delicate point is that the a posteriori technique gives an existence statement, but does not bound the error between ordered eigenpairs. We overcome this difficulty using the absolute Weyl theorem for generalized symmetric eigenvalue problems, see [21]. We conclude that (κ_1, u_1) can be estimated with the rate $N^{-(s+1)/(2s+3)}$.

Because the volatility estimator relies on the first derivative of the eigenfunction the statistical problem is ill-posed of degree one, deteriorating the rate to $N^{-s/(2s+3)}$. For the drift estimator we need the second derivative, adding a degree of ill-posedness. At the same time the regularity of b is smaller such that the rate becomes $N^{-(s-1)/(2s+3)} = N^{-(s-1)/(2(s-1)+5)}$. Compared to Gobet et al. [14], the same rates can thus be achieved with random sampling times (with unknown sampling distribution) than with equidistant low frequent observations. In fact, the convergence rates are optimal in the minmax sense:

Theorem 8. *Grant Assumption 1 for an arbitrary $\gamma \in \Gamma$ admitting a bounded Lebesgue density at the origin. Grant Assumption 3 for some $s > 1$. For $0 < a < b < 1$ it holds*

$$\begin{aligned} \inf_{\bar{\sigma}} \sup_{(\sigma^2, b) \in \Theta_s} \mathbb{E}_{\sigma, b, \gamma} [\|\bar{\sigma}^2 - \sigma^2\|_{L^2([a, b])}^2] &\gtrsim N^{-2s/(2s+3)}, \\ \inf_{\bar{b}} \sup_{(\sigma^2, b) \in \Theta_s} \mathbb{E}_{\sigma, b, \gamma} [\|\bar{b} - b\|_{L^2([a, b])}^2] &\gtrsim N^{-2(s-1)/(2s+3)}, \end{aligned}$$

where the infimum is taken over all estimators, i.e. measurable functions, $\bar{\sigma}$ and \bar{b} , respectively.

The proof of the lower bounds for observations sampled at random times follows the same strategy as for low frequency observations in [14]. Constructing alternatives that admit the same invariant measure, proving the lower bound is reduced to a testing problem by Assouad's lemma, see Tsybakov [32, Sect. 2.7.2]. The Kullback-Leibler distance between the distributions of two alternatives can then be bounded in terms of the L^2 -distance between the kernels of the corresponding operators R from (5), which is finally accomplished using Hilbert-Schmidt norm estimates and the explicit form of the inverse of the generator.

5 Adaptive estimation

The optimal choice of the projection level crucially depends on the unknown smoothness s . In this section, we construct a completely data driven estimation procedure adapting to the Sobolev regularity of σ^2 and b . We focus on the volatility estimator, noting that the methodology should extend to the drift estimation without additional theoretical problems. We adopt the general adaption principle by Lepskiï [20].

The aim is to chose the optimal projection level from the set

$$\mathcal{J}_N := [J_{min}, J_{max}] \cap \mathbb{N} \quad \text{with} \quad 2^{J_{min}} \sim \log N, \quad 2^{J_{max}} \sim \frac{N}{(\log N)^2 \log \log N}$$

For any $J \in \mathcal{J}_N$ we define

$$s_J^2 := \Lambda^2 2^{3J} \frac{\log \log N}{N} \quad (14)$$

for some appropriate constant $\Lambda > 0$ depending on d, D as well as I, α (but not on s) from the Assumptions 1 and 3. The quantity s_J is an upper bound for the stochastic error of $\hat{\sigma}_J^2$, cf. Corollary 24. The adaptive estimator is defined by

$$\tilde{\sigma}^2 := \hat{\sigma}_{\hat{J}}^2 \quad \text{with} \quad \hat{J} := \min \{ J \in \mathcal{J}_N : \forall K \geq J, K \in \mathcal{J}_N \|\hat{\sigma}_K^2 - \hat{\sigma}_J^2\|_{L^2([a,b])} \leq s_K \}.$$

Heuristically, \hat{J} is the smallest projection level for which the stochastic error still dominates the bias.

Our main result for the adaptive estimation shows that the estimator $\tilde{\sigma}^2$ achieves the optimal convergence rate up to an additional $\log \log N$ factor.

Theorem 9. *Grant Assumptions 1 and define $\Gamma_0 := \{\gamma \in \Gamma : \mathbb{E}_\gamma[\tau^{-1/2}] \leq D\}$. Let Assumption 3 be fulfilled for some $s > 5/2$. Let $0 < a < b < 1$. Then there exists for every $\varepsilon > 0$ some $C > 0$ such that, for N sufficiently large, we have*

$$\sup_{(\sigma, b, \gamma) \in \Theta_s \times \Gamma_0} \mathbb{P}_{\sigma, b, \gamma} \left(\|\tilde{\sigma}^2 - \sigma^2\|_{L^2([a,b])}^2 > C \left(\frac{\log \log N}{N} \right)^{2s/(2s+3)} \right) < \varepsilon.$$

The proof of this theorem is postponed to Section 9. It relies on a concentration inequality for the Markov chain $(X_{\tau_n})_{n \geq 0}$, see Proposition 23 as well as Nickl and Söhl [23, Section 3]. For the latter we need the additional assumption on γ allowing for a uniform bound on the transition density of the time-changed diffusion process. Up to the concentration result, the proof relies on the standard arguments for the Lepski method.

6 Numerical example

In this section, we present numerical results for the volatility estimation. Throughout the chapter, we consider a diffusion process X with linear mean reverting drift $b(x) = 0.2 - 0.4x$, quadratic squared volatility function $\sigma^2(x) = 0.4 - (x - 0.5)^2$ and two reflecting barriers at 0 and 1. The sample paths were generated using Euler-Maruyama scheme with time step size 0.001 and reflection after each step.

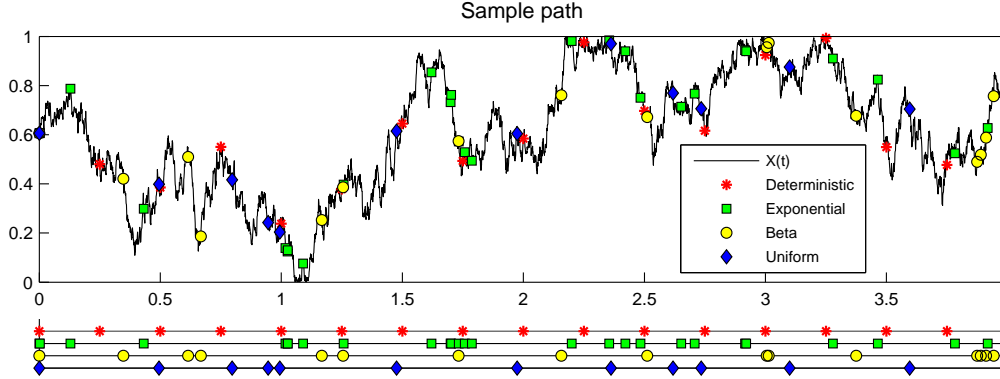


Figure 1: Sample path of the process X for $0 \leq t \leq 4$ with marked observations from different sampling distributions.

Sample Size Distribution	Oracle projection level			Adaptive estimator		
	4 000	12 000	20 000	4 000	12 000	20 000
Deterministic	0.0233	0.0155	0.0123	0.0318	0.0214	0.0130
Uniform	0.0258	0.0168	0.0134	0.0341	0.0221	0.0139
Exponential	0.0282	0.0177	0.0141	0.0362	0.0231	0.0148
Beta	0.0296	0.0211	0.0179	0.0432	0.0255	0.0178

Table 1: Root mean integrated squared error for volatility estimation on $[0.1, 0.9]$ based on 1000 Monte Carlo iterations.

For $\Delta = 0.25$ we compare the estimation error for four different sampling distributions of quite different shapes: the case of equidistant observations with frequency Δ^{-1} , the uniform distribution on the interval $[0, 2\Delta]$, the symmetric Beta(0.2, 0.2) distribution rescaled to the interval $[0, 2\Delta]$ and finally, the exponential distribution with intensity Δ^{-1} . Note that all considered distributions have mean Δ , Uniform and Beta distribution have the same compact support $[0, 2\Delta]$ and together with exponential distribution they allow for arbitrary small sampling distances. Figure 1 depicts a fragment of a simulated trajectory of the diffusion together with the observations from different sampling schemes.

To construct the approximation spaces, we used the Fourier orthogonal cosines basis i.e.

$$V_J = \overline{\text{span}}\{\sqrt{2}\cos(j\pi x) : 0 \leq j \leq J\},$$

cf. Remark 5. We compare an oracle choice of the projection level with the adaptive estimator. As target interval we choose $[0.1, 0.9]$.

In Table 1 we compare the oracle and adaptive root mean integrated squared error (RMISE) for volatility estimation on the interval $[0.1, 0.9]$, obtained by a Monte Carlo simulation with 1000 iterations. The oracle projection level J is stable with respect to the sampling distribution and surprisingly small, taking values 2 for $N = 4\,000$ and 4 for $N = 12\,000$ and $N = 20\,000$ across all distributions, with the exception of Beta with sample size $N = 12\,000$, when it equals 2. For the adaptive estimation we chose the constant Λ in (14) equal to 0.01.

Relative to $\|\sigma^2\|_{L^2([0.1, 0.9])} \approx 0.31$ the error of the oracle decreases from approximately 10% for sample size $N = 4\,000$ to 5% for $N = 20\,000$. In particular for large sample sized the error of the adaptive procedure is fairly close to the oracle error. The errors are quite stable across sampling distributions as the estimator, where the deterministic sampling allows for the smallest error and the Beta distribution generates the largest errors. The latter is not surprising because the Beta distribution is chosen in a way that yields a strong clustering of the observations.

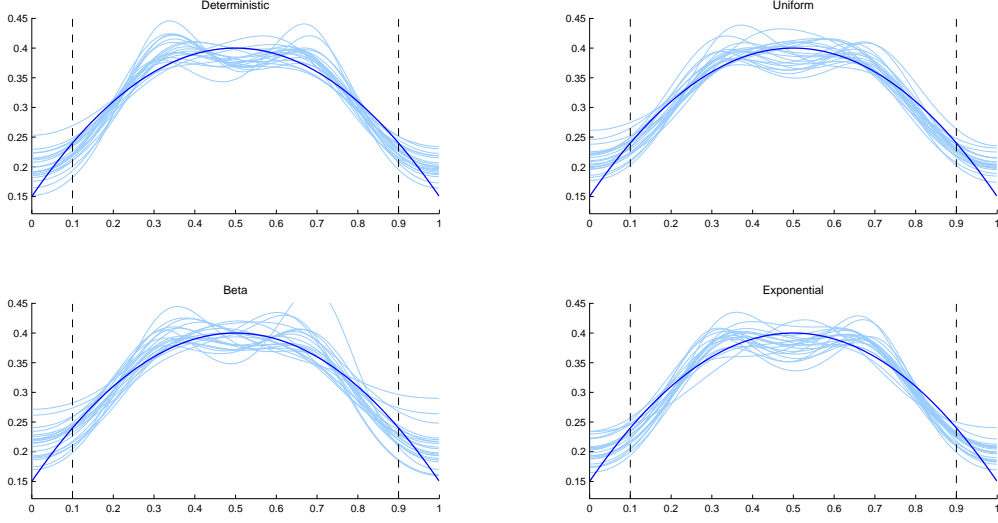


Figure 2: Estimated volatility functions using adapted estimator for 20 independent trajectories of the diffusion and four different sampling distributions with sample size $N = 20\,000$.

For 20 independent paths and sample size $N = 20\,000$ the resulting adaptive volatility estimators are shown in Figure 2. While the estimators behave nicely in the interior of the interval, the boundary problem outside the interval $[0.1, 0.9]$ is clearly visible. Again we see that the estimation for the Beta sampling distribution is the worst.

In the misspecified case where the randomness of the observation times is ignored, the RMISE of the low-frequency estimator designed for equidistant observations with Δ set to the average observation distance is four times larger than the error of our method in our simulations.

7 Proofs of the upper bounds

Throughout we take Assumptions 1 and 3 for granted.

7.1 Spectral properties of the generalized transition operator R

Recall that u_1 is the eigenfunction corresponding to the biggest negative eigenvalue v_1 of the generator L , normalized in $L^2([0, 1])$. By [14, Proposition 6.5] u_1 can be chosen to be increasing and for any $0 < a < b < 1$ there exists a positive constant $c_{a,b} > 0$ such that

$$\inf_{(\sigma,b) \in \Theta_s} \inf_{x \in [a,b]} u_1'(x) > c_{a,b}. \quad (15)$$

By Lemma 6.1 in [14] the family of generators $\{L_{\sigma,b} : (\sigma,b) \in \Theta_s\}$ has a uniform spectral gap on Θ_s meaning that there is a constant $s_0 > 0$ such that

$$\inf_{(\sigma,b) \in \Theta_s} \inf_{i \neq 1} |v_i - v_1| = \inf_{(\sigma,b) \in \Theta_s} \{|v_1|, |v_2 - v_1|\} \geq s_0. \quad (16)$$

Moreover the eigenvalues v_k satisfy uniformly on Θ_s

$$C_1 k^2 \leq -v_k \leq C_2 k^2, \quad (17)$$

for constants $0 < C_1 < C_2$, while corresponding eigenfunctions u_k belong to the Sobolev space H^{s+1} fulfilling

$$\|u_k\|_{H^{s+1}} \lesssim (1 \vee |v_k|)^{\lceil s \rceil}. \quad (18)$$

As announced in Lemma (4) these bounds transfer uniformly to the operator R .

Proof of Lemma 4. For convenience we define $m := \min I > 0$ and $M := \max I$. By the definition of R and the uniform bounds on the eigenvalues v_k of L in (17), we have

$$\kappa_k = \mathcal{L}_\gamma(-v_k) = \int_0^\infty e^{tv_k} \gamma(dt) \geq \int_0^\infty e^{-tC_2 k^2} \gamma(dt) \geq \alpha e^{-MC_2 k^2} \quad \text{for } k \geq 1.$$

The spectral gap of the operator R equals $\min \{1 - \kappa_1, \kappa_1 - \kappa_2\}$. Due to (16), we have

$$\begin{aligned} \kappa_1 - \kappa_2 &= \int_0^\infty (e^{tv_1} - e^{tv_2}) \gamma(dt) = \int_0^\infty e^{tv_2} (e^{t(v_1-v_2)} - 1) \gamma(dt) \\ &\geq \int_0^\infty e^{-4tC_2} (e^{ts_0} - 1) \gamma(dt) \geq \alpha e^{-4MC_2} (e^{ms_0} - 1). \end{aligned}$$

Similarly $1 - \kappa_1 = \int_0^\infty (1 - e^{tv_1}) \gamma(dt) \geq \int_0^\infty (1 - e^{-tC_1}) \gamma(dt) \geq \alpha(1 - e^{-mC_1})$. \square

7.2 Consequences of the mixing property

First we establish general bounds for the variance of integrals with respect to the empirical measure which are due to the mixing behavior of the sequence $(X_{\tau_k})_k$. The following Lemma is a straightforward generalization of [14, Lemma 6.2]. Since this is the key result to bound the stochastic error, we give the proof to keep the paper self-contained.

Lemma 10. *For bounded $H_1, H_2 \in L^2([0, 1])$ we have the following two variance estimates:*

$$\begin{aligned} \text{Var}_{\sigma, b, \gamma} \left[\frac{1}{N} \sum_{n=1}^N H_1(X_{\tau_n}) \right] &\lesssim N^{-1} \mathbb{E}_{\sigma, b, \gamma} [H_1^2(X_0)], \\ \text{Var}_{\sigma, b, \gamma} \left[\frac{1}{N} \sum_{n=0}^{N-1} H_1(X_{\tau_n}) H_2(X_{\tau_{n+1}}) \right] &\lesssim N^{-1} \mathbb{E}_{\sigma, b, \gamma} [H_1^2(X_0) H_2^2(X_{\tau_1})]. \end{aligned}$$

Proof. Denote $f(X_{\tau_n}) = H_1(X_{\tau_n}) - \mathbb{E}_{\sigma, b, \gamma} [H_1(X_{\tau_n})]$. Consider $m \geq n$ and let $k = m - n$. Since process X is stationary and has a uniform spectral gap $\|R^k f\|_{L^2(\mu)} \leq \|f\|_{L^2(\mu)} \mathcal{L}_\gamma^k(s_0)$ holds for every function f that is $L^2(\mu)$ -orthogonal to constants. Arguing analogously as in the proof of Lemma 4 we obtain $\sup_{\gamma \in \Gamma} \mathcal{L}_\gamma(s_0) < 1$. Hence, by the Cauchy-Schwarz inequality,

$$\begin{aligned} \mathbb{E}_{\sigma, b, \gamma} [f(X_{\tau_n}) f(X_{\tau_m})] &= \mathbb{E}_{\sigma, b, \gamma} [f(X_{\tau_n}) \mathbb{E}_{\sigma, b, \gamma} [f(X_{\tau_{n+k}}) | X_{\tau_n}]] \\ &= \langle f, R^k f \rangle_\mu \leq \|f\|_{L^2(\mu)}^2 \mathcal{L}_\gamma^k(s_0). \end{aligned}$$

Since $\|f\|_{L^2(\mu)}^2 = \text{Var}_{\sigma, b, \gamma} [H_1(X_0)] \leq \mathbb{E}_{\sigma, b, \gamma} [H_1^2(X_0)]$ and

$$\text{Var}_{\sigma, b, \gamma} \left[\sum_{n=1}^N H_1(X_{\tau_n}) \right] = \sum_{n, m=1}^N \mathbb{E}_{\sigma, b, \gamma} [f(X_{\tau_n}) f(X_{\tau_m})] \leq \|f\|_{L^2(\mu)}^2 \sum_{n, m=1}^N \mathcal{L}_\gamma^{|n-m|}(s_0)$$

to prove the first inequality we just have to show that $\sum_{n, m=1}^N \mathcal{L}_\gamma^{|n-m|}(s_0) \lesssim N$. This easily follows from the formula for the sum of finite geometric series.

To prove the second inequality, first note that

$$\text{Var}_{\sigma, b, \gamma} \left[\frac{1}{N} \sum_{n=0}^{N-1} H_1(X_{\tau_n}) H_2(X_{\tau_{n+1}}) \right] \leq \frac{1}{N^2} \mathbb{E}_{\sigma, b, \gamma} \left[\sum_{n, m=0}^{N-1} H_1(X_{\tau_n}) H_2(X_{\tau_{n+1}}) H_1(X_{\tau_m}) H_2(X_{\tau_{m+1}}) \right].$$

$$= \frac{1}{N^2} \mathbb{E}_{\sigma, b, \gamma} \left[\sum_{n, m=0}^{N-1} H_1(X_{\tau_n}) H_2(X_{\tau_{n+1}}) H_1(X_{\tau_m}) H_2(X_{\tau_{m+1}}) \right] - \langle H_1, RH_2 \rangle_\mu^2.$$

Since the sum of diagonal terms equals $N^{-1} \mathbb{E}_{\sigma, b, \gamma} [H_1^2(X_0) H_2^2(X_{\tau_1})]$, it does not exceed the claimed upper bound. The sum of the other terms equals

$$\frac{1}{N^2} \sum_{\substack{n, m=0 \\ n \neq m}}^{N-1} \langle H_2 \cdot (RH_1), R^{|n-m|-1} (H_1 \cdot (RH_2) - \langle H_1, RH_2 \rangle_\mu) \rangle_\mu - \frac{1}{N} \underbrace{\langle H_1, RH_2 \rangle_\mu^2}_{\lesssim N^{-1} \mathbb{E}_{\sigma, b, \gamma} [H_1^2(X_0) H_2^2(X_{\tau_1})]}.$$

Using the spectral gap of the operator R together with the Cauchy-Schwarz inequality, we obtain that

$$\left\| R^{|n-m|-1} (H_1 \cdot (RH_2) - \langle H_1, RH_2 \rangle_\mu) \right\|_{L^2(\mu)} \lesssim \|H_1 \cdot (RH_2)\|_{L^2(\mu)} \mathcal{L}_\gamma^{|n-m|-1}(s_0).$$

Consequently, using again Cauchy-Schwarz and the formula for the sum of finite geometric series, we can bound the considered variance by

$$\begin{aligned} & \frac{1}{N^2} \sum_{\substack{n, m=0 \\ n \neq m}}^{N-1} \|H_2 \cdot (RH_1)\|_{L^2(\mu)} \|H_1 \cdot (RH_2)\|_{L^2(\mu)} \mathcal{L}_\gamma^{|n-m|-1}(s_0) \\ & \lesssim \frac{1}{N} \|H_2 \cdot (RH_1)\|_{L^2(\mu)} \|H_1 \cdot (RH_2)\|_{L^2(\mu)} \\ & \lesssim \frac{1}{N} \mathbb{E}_{\sigma, b, \gamma} [H_2^2(X_0) H_1^2(X_{\tau_1})]^{1/2} \mathbb{E}_{\sigma, b, \gamma} [H_1^2(X_0) H_2^2(X_{\tau_1})]^{1/2} \\ & = \frac{1}{N} \mathbb{E}_{\sigma, b, \gamma} [H_2^2(X_0) H_1^2(X_{\tau_1})]. \end{aligned} \quad \square$$

The first consequence of the previous result is the following bound for the risk of the estimator of the invariant measure.

Proposition 11. *Under Assumption 3 it holds*

$$\mathbb{E}_{\sigma, b, \gamma} \left[\|\mu - \hat{\mu}_J\|_{L^2}^2 \right] \lesssim N^{-2Js} + N^{-1} 2^J. \quad (19)$$

Furthermore if we choose $2^J \sim N^{1/(2s+3)}$ the event $\mathcal{T}_0 = \{\forall x \in [0, 1] \inf \mu/2 \leq \hat{\mu}_J(x) \leq 2 \sup \mu\}$ satisfies $\mathbb{P}_{\sigma, b, \gamma}(\Omega \setminus \mathcal{T}_0) \lesssim N^{-\frac{2s}{2s+3}}$.

Proof. The explicit formula (2) for μ shows that $\|\mu\|_{H^s}$ is uniformly bounded over Θ_s . Jackson's inequality yields

$$\|(I - \pi_J)\mu\|_{L^2}^2 \lesssim 2^{-2Js}.$$

Using Lemma 10, we obtain

$$\begin{aligned} \mathbb{E}_{\sigma, b, \gamma} [\|\pi_J \mu - \hat{\mu}_J\|_{L^2}^2] &= \sum_{|\lambda| \leq J} \mathbb{E}_{\sigma, b, \gamma} [\langle \psi_\lambda, \mu - \mu_N \rangle^2] = \sum_{|\lambda| \leq J} \text{Var}_{\sigma, b, \gamma} [\langle \psi_\lambda, \mu_N \rangle] \\ &\lesssim N^{-1} \sum_{|\lambda| \leq J} \mathbb{E}_{\sigma, b, \gamma} [\psi_\lambda^2(X_0)] \lesssim 2^J N^{-1} \end{aligned}$$

and (19) follows by the triangle inequality. Furthermore, by Jackson's inequality,

$$\begin{aligned} \sup_{x \in [0, 1]} \pi_J \mu(x) &\leq \|\mu\|_\infty + \|(I - \pi_J)\mu\|_\infty \lesssim \|\mu\|_{H^1} + \|(I - \pi_J)\mu\|_{H^1} \lesssim 1 + 2^{-J(s-1)} \\ \inf_{x \in [0, 1]} \pi_J \mu(x) &\geq \inf_{x \in [0, 1]} \mu(x) - \|(I - \pi_J)\mu\|_\infty \gtrsim 1 - 2^{-J(s-1)}. \end{aligned}$$

Hence, for J large enough, $\pi_J \mu$ is bounded by $\frac{3}{4} \inf \mu$ from below and $\frac{3}{2} \sup \mu$ from above. Consequently, $\hat{\mu}_J(x)$ lies in $[\frac{1}{2} \inf \mu, 2 \sup \mu]$ if $\|\hat{\mu}_J - \pi_J \mu\|_\infty$ is small enough. For a given constant $C > 0$, Bernstein's inequality shows

$$\begin{aligned} \mathbb{P}_{\sigma,b,\gamma}(\|\hat{\mu}_J - \pi_J \mu\|_\infty > C) &\leq C^{-2} \mathbb{E}_{\sigma,b,\gamma}[\|\pi_J \mu - \hat{\mu}_J\|_\infty^2] \lesssim \mathbb{E}_{\sigma,b,\gamma}[\|\pi_J \mu - \hat{\mu}_J\|_{H^1}^2] \\ &\lesssim 2^{2J} \mathbb{E}_{\sigma,b,\gamma}[\|\pi_J \mu - \hat{\mu}_J\|_{L^2}^2] \lesssim N^{-\frac{2s}{2s+3}}. \end{aligned} \quad \square$$

7.3 Analysis of the projection error

Denote by $(\kappa_{J,i}, u_{J,i})$, $i = 0, 1, 2, \dots, \dim V_J - 1$, the eigenpairs of the operator $\pi_J^\mu R \pi_J^\mu$ ordered decreasingly with respect to the eigenvalues. Note that $(\kappa_{J,i}, u_{J,i})$ are solutions of the eigenvalue problem for the operator R restricted to the finite approximation spaces V_J on $L^2(\mu)$:

$$\langle Ru_{J,i}, v \rangle_\mu = \kappa_{J,i} \langle u_{J,i}, v \rangle_\mu, \text{ for every } v \in V_J. \quad (20)$$

Take $u_{J,i}$ normalized in the L^2 norm. Since $\pi_J^\mu R \pi_J^\mu$ is a positive definite self-adjoint operator on $L^2(\mu)$ with $\|\pi_J^\mu R \pi_J^\mu\|_{L^2(\mu)} \leq 1$ we have $0 < \kappa_{J,i} \leq 1$.

Proposition 12. *For sufficiently large J it holds uniformly on Θ_s*

$$|\kappa_{J,1} - \kappa_1| + \|u_{J,1} - u_1\|_{H^1} \lesssim 2^{-Js}.$$

Proof. It suffices to show that $|\kappa_{J,1} - \kappa_1| + \|u_{J,1} - u_1\|_{L^2} \lesssim 2^{-J(s+1)}$. Indeed, by Jackson's and Bernstein's inequalities

$$\begin{aligned} \|u_{J,1} - u_1\|_{H^1} &\leq \|u_{J,1} - \pi_J u_1\|_{H^1} + \|(I - \pi_J) u_1\|_{H^1} \lesssim 2^J \|u_{J,1} - \pi_J u_1\|_{L^2} + \|(I - \pi_J) u_1\|_{H^1} \\ &\lesssim 2^J \|u_{J,1} - u_1\|_{L^2} + 2^J \|(I - \pi_J) u_1\|_{L^2} + \|(I - \pi_J) u_1\|_{H^1} \\ &\lesssim 2^J \|u_{J,1} - u_1\|_{L^2} + 2^{-Js} \end{aligned}$$

where we used the upper bound (18).

Recall that R is a compact self-adjoint positive-definite operator on $L^2(\mu)$. Furthermore

$$\begin{aligned} \|(I - \pi_J^\mu) u_1\|_{L^2(\mu)} &\lesssim \|(I - \pi_J^\mu) (I - \pi_J) u_1\|_{L^2} \lesssim \|(I - \pi_J) u_1\|_{L^2} \\ &\lesssim 2^{-J(s+1)} \|u_1\|_{H^{s+1}} \lesssim 2^{-J(s+1)}. \end{aligned}$$

Consequently, since by Lemma 4 operator R has a uniform spectral gap inequality

$$\|(I - \pi_J^\mu) u_1\|_{L^2(\mu)} \leq \frac{\kappa_1 - \kappa_2}{4\kappa_1}$$

holds for J large enough. It follows that we can use Theorem 25 obtaining

$$|\kappa_{J,1} - \kappa_1| + \left\| \frac{u_{J,1}}{\|u_{J,1}\|_{L^2(\mu)}} - \frac{u_1}{\|u_1\|_{L^2(\mu)}} \right\|_{L^2(\mu)} \lesssim 2^{-J(s+1)}.$$

The claim follows since $\|u_{J,1} - u_1\|_{L^2} \lesssim \left\| \frac{u_{J,1}}{\|u_{J,1}\|_{L^2(\mu)}} - \frac{u_1}{\|u_1\|_{L^2(\mu)}} \right\|_{L^2(\mu)}$ by the equivalence of norms $\|\cdot\|_{L^2}$ and $\|\cdot\|_{L^2(\mu)}$. \square

Corollary 13. *Projected operators $\pi_J^\mu R \pi_J^\mu$ have a uniform spectral gap, i.e. there exists $s_1 > 0$ such that*

$$\min\{|\kappa_{J,1}|, |\kappa_{J,2} - \kappa_{J,1}|\} \geq s_1$$

for every J large enough.

Proof. Follows from the proof of Theorem 25. \square

7.4 Analysis of the stochastic error

Define the operator $R_J : V_J \rightarrow V_J$ as the restriction of the operator $\pi_J^\mu R \pi_J^\mu$ to the finite dimensional Hilbert space V_J . Recall that the operator G_J was defined by the Gram matrix of the inner product $\langle \cdot, \cdot \rangle_\mu$, i.e. for $v \in V_J$ we have $\langle v, G_J v \rangle = \langle v, v \rangle_\mu$. Note that by (20)

$$R_J u_{J,i} = \kappa_{J,i} G_J u_{J,i}, \quad (21)$$

hence $(\kappa_{J,i}, u_{J,i})$ are solutions of generalized symmetric eigenvalue problem for R_J, G_J . When matrix \hat{G}_J is invertible the corresponding generalized eigenvalue problem for \hat{G}_J, \hat{R}_J , namely

$$\hat{R}_J \hat{u}_{J,i} = \hat{\kappa}_{J,i} \hat{G}_J \hat{u}_{J,i} \quad (22)$$

has $\dim V_J$ solutions that we denote by $(\hat{\kappa}_{J,i}, \hat{u}_{J,i})$, $i = 0, 1, \dots, \dim V_J - 1$. Recall that the eigenfunctions $\hat{u}_{J,i}$ are normalized in $L^2[0, 1]$.

In this subsection we want to bound the expected error between $(\kappa_{J,1}, u_{J,1})$ and $(\hat{\kappa}_{J,1}, \hat{u}_{J,1})$. From the general theory of a posteriori error bound techniques for generalized symmetric eigenvalue problems (see Section A.2) we know that the error between the eigenpairs can be controlled by the norm of the residual vectors:

$$r = (\hat{R}_J - R_J)u_{J,1} + \kappa_{J,1}(G_J - \hat{G}_J)u_{J,1} \text{ or } r^* = (R_J - \hat{R}_J)\hat{u}_{J,1} + \hat{\kappa}_{J,1}(\hat{G}_J - G_J)\hat{u}_{J,1}.$$

Since the eigenpair $(\hat{\kappa}_{J,1}, \hat{u}_{J,1})$ of the problem (22) is random and depends on operators \hat{R}_J and \hat{G}_J it is easier to analyze the norm of the vector r rather than r^* (cf. Lemmas 14 and 15 where v is a deterministic function). Consequently in the following we refer to r as the residual vector. In the notation of Section A.2 we treat the deterministic problem (21) as a perturbed approximation of the data dependent problem (22).

Lemma 14. *For any $v \in V_J$ we have, uniformly on $\Theta_s \times \Gamma$,*

$$\mathbb{E}_{\sigma,b,\gamma} \left[\|(\hat{G}_J - G_J)v\|_{L^2}^2 \right] \lesssim N^{-1} 2^J \|v\|_{L^2}^2.$$

Proof. Given Lemma 10, the proof is a straight forward estimate analogously to [14, Lemma 4.8]. \square

Now, we are ready to prove Lemma 6:

Proof of Lemma 6. A standard Neumann series argument shows that \hat{G}_J is invertible on \mathcal{T}_1 with $\|\hat{G}_J^{-1}\|_{L^2} \leq 2\|G_J^{-1}\|_{L^2}$. Since the invariant density μ has a positive lower bound uniformly on Θ_s , for any $v \in V_J$ we have

$$\langle v, G_J v \rangle = \langle v, v \rangle_\mu = \|v\|_{L^2(\mu)}^2 \gtrsim \|v\|_{L^2}^2.$$

Hence the smallest eigenvalue of the operator G_J is uniformly separated from zero. This implies that G_J^{-1} is uniformly bounded in the operator norm. The classical Hilbert-Schmidt norm inequality yields

$$\|\hat{G}_J - G_J\|_{L^2}^2 \leq \sum_{|\lambda| \leq J} \|(\hat{G}_J - G_J)\psi_\lambda\|_{L^2}^2.$$

Consequently, by Lemma 14, $\mathbb{E}_{\sigma,b,\gamma} [\|\hat{G}_J - G_J\|_{L^2}^2] \lesssim N^{-1} 2^{2J}$ and $\mathbb{P}_{\sigma,b,\gamma}(\Omega \setminus \mathcal{T}_1) \leq N^{-1} 2^{2J}$ follows from Chebyshev's inequality. \square

Lemma 15. *For any $v \in V_J$ we have, uniformly on $\Theta_s \times \Gamma$,*

$$\mathbb{E}_{\sigma,b,\gamma} \left[\|(\hat{R}_J - R_J)v\|_{L^2}^2 \right] \lesssim N^{-1} 2^J \|v\|_{L^2}^2.$$

Proof. By Lemma 10 we obtain

$$\begin{aligned}
\mathbb{E}_{\sigma,b,\gamma} \left[\|\widehat{R}_J - R_J\|_{L^2}^2 \right] &= \sum_{|\lambda| \leq J} \text{Var}_{\sigma,b,\gamma} \left[\frac{1}{N} \sum_{n=0}^{N-1} \psi_\lambda(X_{\tau_n}) v(X_{\tau_n}) \right] \\
&\lesssim \sum_{|\lambda| \leq J} N^{-1} \mathbb{E}_{\sigma,b,\gamma} [\psi_\lambda^2(X_{\tau_1}) v^2(X_0)] \\
&\lesssim N^{-1} \left\| \sum_{|\lambda| \leq J} \psi_\lambda^2 \right\|_\infty \mathbb{E}_{\sigma,b,\gamma} [v^2(X_0)] \\
&\lesssim N^{-1} 2^J \|v^2\|_{L^2(\mu)}^2. \quad \square
\end{aligned}$$

Corollary 16. *We have, uniformly on $\Theta_s \times \Gamma$, the following bound on the norm of the residual vector $r = (\widehat{R}_J - R_J)u_{J,1} + \kappa_{J,1}(G_J - \widehat{G}_J)u_{J,1}$*

$$\mathbb{E}_{\sigma,b,\gamma} \left[\|r\|_{L^2}^2 \right] \lesssim N^{-1} 2^J.$$

Proof. Note that from Proposition 12 we know that, for J big enough, the eigenvalue $\kappa_{J,1}$ is uniformly bounded. Consequently

$$\mathbb{E}_{\sigma,b,\gamma} \left[\|r\|_{L^2}^2 \right] \lesssim \mathbb{E}_{\sigma,b,\gamma} \left[\|(\widehat{R}_J - R_J)u_{J,1}^J\|_{L^2}^2 \right] + \mathbb{E}_{\sigma,b,\gamma} \left[\|(\widehat{G}_J - G_J)u_{J,1}^J\|_{L^2}^2 \right] \lesssim N^{-1} 2^J$$

by Lemmas 14 and 15. \square

Proposition 17. *On the event \mathcal{T}_1 the eigenpair $(\widehat{\kappa}_{J,1}, \widehat{u}_{J,1})$ is the biggest nontrivial eigenpair of the matrix $\widehat{G}_J^{-1} \widehat{R}_J$. Furthermore there exists a set $\mathcal{T}_2 \subset \mathcal{T}_1$ such that*

$$\mathbb{P}_{\sigma,b,\gamma}(\Omega \setminus \mathcal{T}_2) \lesssim N^{-1} 2^{3J}$$

and

$$\mathbb{E}_{\sigma,b,\gamma} \left[\mathbf{1}_{\mathcal{T}_2} \cdot \left(|\kappa_{J,1} - \widehat{\kappa}_{J,1}|^2 + \|u_{J,1} - \widehat{u}_{J,1}\|_{L^2}^2 \right) \right] \lesssim N^{-1} 2^J$$

holds uniformly on Θ_s .

Proof. By Theorem 26 there exists some $0 \leq i_0 \leq \dim V_J - 1$ such that the eigenpair $(\widehat{\kappa}_{J,i_0}, \widehat{u}_{J,i_0})$ of the problem (22) satisfies

$$\begin{aligned}
|\kappa_{J,1} - \widehat{\kappa}_{J,i_0}| &\leq \|\widehat{G}_J^{-1}\|_{L^2} \|r\|_{L^2}, \\
\|u_{J,1} - \widehat{u}_{J,i_0}\|_{L^2} &\leq \frac{2\sqrt{2}}{\delta(\widehat{\kappa}_{J,i_0})} \|\widehat{G}_J\|_{L^2}^{1/2} \|\widehat{G}_J^{-1}\|_{L^2}^{3/2} \|r\|_{L^2},
\end{aligned}$$

where $\delta(\widehat{\kappa}_{J,i_0}) = \min_{j \neq i_0} \{|\widehat{\kappa}_{J,j} - \widehat{\kappa}_{J,i_0}|\}$ is the isolation distance of the eigenvalues $\widehat{\kappa}_{J,i_0}$ and $\kappa_{J,1}$. Let s_1 be the uniform spectral gap of operators R_J (see Corollary 13). Define \mathcal{T}_2 as the subset of \mathcal{T}_1 for which $i_0 = 1$ and $\delta(\widehat{\kappa}_{J,1}) \geq \frac{1}{2}s_1$. Since $\|\widehat{G}_J^{-1}\|_{L^2}$ and $\|\widehat{G}_J\|_{L^2}$ are uniformly bounded on the event \mathcal{T}_1 and $\mathbb{E}_{\sigma,b,\gamma}[\|r\|_{L^2}^2] \lesssim N^{-1} 2^J$ the desired error bound holds when we restrict to the event \mathcal{T}_2 .

To finish the proof we must show that $\mathbb{P}_{\sigma,b,\gamma}(\Omega \setminus \mathcal{T}_2) \lesssim N^{-1} 2^{3J}$. Denote

$$\mathcal{T}_2 = \underbrace{\mathcal{T}_1 \cap \{i_0 = 1\}}_{\mathcal{T}_{2,1}} \cap \underbrace{\{\delta(\widehat{\kappa}_{J,1}) \geq s_1/2\}}_{\mathcal{T}_{2,2}}.$$

First, using the absolute Weyl theorem (Theorem 27) we observe that for any $0 \leq j \leq \dim V_J - 1$

$$\mathbb{E}_{\sigma,b,\gamma} \left[\mathbf{1}_{\mathcal{T}_1} \cdot |\kappa_{J,j} - \widehat{\kappa}_{J,j}|^2 \right] \leq \mathbb{E}_{\sigma,b,\gamma} \left[\mathbf{1}_{\mathcal{T}_1} \cdot \|\widehat{G}_J^{-1}\|_{L^2}^2 \|(R_J - \widehat{R}_J) - \kappa_{J,j}(G_J - \widehat{G}_J)\|_{L^2}^2 \right]$$

$$\begin{aligned}
&\lesssim \mathbb{E}_{\sigma,b,\gamma} \left[\mathbf{1}_{\mathcal{T}_1} \cdot \|R_J - \widehat{R}_J\|_{L^2}^2 \right] + \kappa_{J,j} \mathbb{E}_{\sigma,b,\gamma} \left[\mathbf{1}_{\mathcal{T}_0} \cdot \|G_J - \widehat{G}_J\|_{L^2}^2 \right] \\
&\lesssim N^{-1} 2^{2J}
\end{aligned}$$

by the classical Hilbert-Schmidt norm inequality. Consequently, using the uniform lower bound on the spectral gap of R_J , we obtain

$$\begin{aligned}
\mathbb{P}_{\sigma,b,\gamma}(\mathcal{T}_1 \setminus \mathcal{T}_{2,1}) &\lesssim \mathbb{E}_{\sigma,b,\gamma} \left[\mathbf{1}_{\mathcal{T}_1 \setminus \mathcal{T}_{2,1}} \cdot |\kappa_{J,2} - \kappa_{J,1}|^2 \right] \\
&\lesssim \mathbb{E}_{\sigma,b,\gamma} \left[\mathbf{1}_{\mathcal{T}_1 \setminus \mathcal{T}_{2,1}} \cdot |\kappa_{J,i_0} - \kappa_{J,1}|^2 \right] \\
&\lesssim \mathbb{E}_{\sigma,b,\gamma} \left[\mathbf{1}_{\mathcal{T}_1 \setminus \mathcal{T}_{2,1}} \cdot |\kappa_{J,i_0} - \widehat{\kappa}_{J,i_0}|^2 \right] + \mathbb{E}_{\sigma,b,\gamma} \left[\mathbf{1}_{\mathcal{T}_1 \setminus \mathcal{T}_{2,1}} \cdot |\widehat{\kappa}_{J,i_0} - \kappa_{J,1}|^2 \right] \\
&\lesssim N^{-1} 2^{2J}.
\end{aligned}$$

Consider now the event $\mathcal{T}_{2,2}$. Since

$$\begin{aligned}
\delta(\widehat{\kappa}_{J,1}) &= \min_{j \neq 1} |\widehat{\kappa}_{J,j} - \kappa_{J,1}| \geq \min_{j \neq 1} \{ |\kappa_{J,j} - \kappa_{J,1}| - |\widehat{\kappa}_{J,j} - \kappa_{J,j}| \} \\
&\geq s_1 - \max_{j \neq 1} \{ |\widehat{\kappa}_{J,j} - \kappa_{J,j}| \},
\end{aligned}$$

we have

$$\begin{aligned}
\mathbb{P}_{\sigma,b,\gamma}(\mathcal{T}_1 \setminus \mathcal{T}_{2,2}) &\leq \mathbb{P}_{\sigma,b,\gamma}(\mathcal{T}_1 \cap \{ \max_{j \neq 1} \{ |\widehat{\kappa}_{J,j} - \kappa_{J,j}| \} \geq s_1/2 \}) \\
&\leq \sum_{1 < j \leq \dim V_J - 1} \mathbb{P}_{\sigma,b,\gamma}(\mathcal{T}_1 \cap \{ |\widehat{\kappa}_{J,j} - \kappa_{J,j}| \geq s_1/2 \}) \\
&\lesssim \sum_{1 < j \leq \dim V_J - 1} \mathbb{E}_{\sigma,b,\gamma} \left[\mathbf{1}_{\mathcal{T}_1} \cdot |\widehat{\kappa}_{J,j} - \kappa_{J,j}|^2 \right] \lesssim N^{-1} 2^{3J}. \quad \square
\end{aligned}$$

7.5 Proof of Theorem 7

From now on we chose $2^J \sim N^{1/(2s+3)}$. Recall that the biggest negative eigenvalue of the infinitesimal generator L is denoted by v_1 which is estimated by $\widehat{v}_{J,1}$ from (10).

Lemma 18. *Choose $2^J \sim N^{1/(2s+3)}$. There is an event $\mathcal{T}_3 \subset \mathcal{T}_2$ satisfying $\mathbb{P}_{\sigma,b,\gamma}(\Omega \setminus \mathcal{T}_3) \lesssim N^{-2s/(2s+3)}$ uniformly on $\Theta_s \times \Gamma$ and*

$$\sup_{(\sigma,b,\gamma) \in \Theta_s \times \Gamma} \mathbb{E}_{\sigma,b,\gamma} [\mathbf{1}_{\mathcal{T}_3} |v_1 - \widehat{v}_{J,1}|^2] \lesssim N^{-\frac{2s}{2s+3}}.$$

In particular we can assume that $|\widehat{v}_{J,1}|$ is uniformly bounded on \mathcal{T}_3 .

Proof. For convenience we denote $m := \min I$, $M := \max I$. On \mathcal{T}_2 we have $\widehat{\kappa}_{J,1} > 0$ and thus $\widehat{\kappa}_{J,1} = \widehat{\mathcal{L}}(-\widehat{v}_{J,1})$.

Step 1: Let us start with a consistency result for $\widehat{v}_{J,1}$. Since $\widehat{\mathcal{L}}$ is non-increasing and continuous, we have for any fixed $\varepsilon \in (0, C_1)$ with C_1 from (17)

$$\mathbb{P}_\gamma(|\widehat{v}_{J,1} - v_1| < \varepsilon) \geq \mathbb{P}_\gamma(\widehat{\mathcal{L}}(-v_1 + \varepsilon) < \widehat{\kappa}_{J,1} < \widehat{\mathcal{L}}(-v_1 - \varepsilon)).$$

Using

$$\delta := \alpha m e^{(v_1 - \varepsilon)M} \leq \inf_{\gamma \in \Gamma} \inf_{|y + v_1| \leq \varepsilon} |\mathcal{L}'_\gamma(y)|, \quad (23)$$

we have $|\mathcal{L}_\gamma(-v_1) - \mathcal{L}_\gamma(-v_1 \pm \varepsilon)| \geq \delta \varepsilon$ uniformly in $\gamma \in \Gamma$ and

$$\mathbb{P}_{\sigma,b,\gamma}(|\widehat{v}_{J,1} - v_1| \geq \varepsilon) \leq \mathbb{P}_{\sigma,b,\gamma}(\kappa_1 - \widehat{\kappa}_{J,1} > \kappa_1 - \widehat{\mathcal{L}}(-v_1 + \varepsilon)) + \mathbb{P}_{\sigma,b,\gamma}(\widehat{\kappa}_{J,1} - \kappa_1 > \widehat{\mathcal{L}}(-v_1 - \varepsilon) - \kappa_1)$$

$$\begin{aligned} &\leq \sum_{y \in \{-\varepsilon, +\varepsilon\}} \mathbb{P}_{\sigma, b, \gamma}(|\hat{\kappa}_{J,1} - \kappa_1| + |\hat{\mathcal{L}}(-v_1 + y) - \mathcal{L}_\gamma(-v_1 + y)| > \delta\varepsilon) \\ &\leq 2\mathbb{P}_{\sigma, b}(|\hat{\kappa}_{J,1} - \kappa_1| > \frac{\delta\varepsilon}{2}) + \sum_{y \in \{-\varepsilon, +\varepsilon\}} \mathbb{P}_\gamma(|\hat{\mathcal{L}}(-v_1 + y) - \mathcal{L}_\gamma(-v_1 + y)| > \frac{\delta\varepsilon}{2}). \end{aligned}$$

By Propositions 12 and 17 and Markov's inequality the first probability is of the order $N^{-2s/(2s+3)}$ if $2^J \sim N^{1/(2s+3)}$. For the estimation error of $\hat{\mathcal{L}}$ Markov's inequality yields for any $y > 0$

$$\begin{aligned} \mathbb{P}_\gamma(|\hat{\mathcal{L}}(y) - \mathcal{L}_\gamma(y)| > \delta\varepsilon/2) &\leq 2(\delta\varepsilon)^{-2} \mathbb{E}_\gamma[|\hat{\mathcal{L}}(y) - \mathcal{L}_\gamma(y)|^2] \\ &= \frac{2}{N\delta^2\varepsilon^2} \text{Var}_\gamma(e^{-y\Delta_1}) \leq \frac{2\mathcal{L}_\gamma(2y)}{N\delta^2\varepsilon^2}. \end{aligned}$$

Therefore,

$$\mathbb{P}_{\sigma, b, \gamma}(|\hat{v}_{J,1} - v_1| \geq \varepsilon) \lesssim N^{-2s/(2s+3)}. \quad (24)$$

Step 2: To determine the rate of $\hat{v}_{J,1}$, we use a Taylor expansion which yields for some intermediate point ξ between $-v_1$ and $-\hat{v}_{J,1}$

$$\hat{\kappa}_{J,1} = \hat{\mathcal{L}}(-\hat{v}_{J,1}) = \hat{\mathcal{L}}(-v_1) + (v_1 - \hat{v}_{J,1})\hat{\mathcal{L}}'(\xi).$$

Since on the other hand we have $\hat{\kappa}_{J,1} = \mathcal{L}_\gamma(-v_1) + \hat{\kappa}_{J,1} - \kappa_1$, we conclude

$$v_1 - \hat{v}_{J,1} = \frac{\mathcal{L}_\gamma(-v_1) - \hat{\mathcal{L}}(-v_1) + \hat{\kappa}_{J,1} - \kappa_1}{\hat{\mathcal{L}}'(\xi)},$$

provided the denominator can be uniformly bounded with high probability. By (24) the event $\mathcal{T}_{3,1} := \{|\hat{v}_{J,1} - v_1| < \varepsilon\}$ has at least the probability $1 - cN^{-2s/(2s+3)}$ for some $c > 0$. On $\mathcal{T}_{3,1}$ we have

$$|\hat{\mathcal{L}}'(\xi)| \geq \inf_{|y+v_1| < \varepsilon} \mathcal{L}'_\gamma(y) - \sup_{|y+v_1| < \varepsilon} |\hat{\mathcal{L}}'(y) - \mathcal{L}'_\gamma(y)|.$$

With δ from (23) we conclude that $|\hat{\mathcal{L}}'(\xi)| \geq \delta/2$ on the event $\mathcal{T}_{3,2} := \{\sup_{y \in [-v_1 - \varepsilon, -v_1 + \varepsilon]} |\hat{\mathcal{L}}'(y) - \mathcal{L}'_\gamma(y)|^2 < \delta/2\}$. Note that in $\mathcal{T}_{3,2}$ we take the supremum of the empirical processes related to $(\Delta_n)_{n=1, \dots, N}$ acting on the function set $\mathcal{F} := \{[0, \infty) \ni x \mapsto xe^{-yx} : y \in [|v_1| - \varepsilon, |v_1| + \varepsilon]\}$. Since \mathcal{F} is the multiplication of the identity map with the transition class $\{e^{-yx} : y > 0\}$, \mathcal{F} is a Vapnik-Červonenkis class and admits the constant envelope function $(|v_1| - \varepsilon)^{-1}e^{-1}$. The empirical process theory (e.g., van der Vaart and Wellner [33], Thm. 2.14.1) yields

$$\mathbb{E}_\gamma \left[\sup_{y \in [-v_1 - \varepsilon, -v_1 + \varepsilon]} |\hat{\mathcal{L}}'(y) - \mathcal{L}'_\gamma(y)|^2 \right] \lesssim \frac{1}{N(|v_1| - \varepsilon)^2}$$

and by Markov's inequality $\mathbb{P}_\gamma(\Omega \setminus \mathcal{T}_{3,2}) \lesssim 1/N$. With $\mathcal{T}_3 := \mathcal{T}_{3,1} \cap \mathcal{T}_{3,2} \cap \mathcal{T}_2$ we finally obtain

$$\begin{aligned} \mathbb{E}_{\sigma, b, \gamma}[\mathbf{1}_{\mathcal{T}_3}|v_1 - \hat{v}_{J,1}|^2] &\leq 2\mathbb{E}_{\sigma, b, \gamma} \left[\mathbf{1}_{\mathcal{T}_3} \frac{|\mathcal{L}_\gamma(-v_1) - \hat{\mathcal{L}}(-v_1)|^2 + |\bar{\kappa}_1 - \kappa_1|^2}{|\hat{\mathcal{L}}'(\xi)|^2} \right] \\ &\lesssim N^{-1} + \mathbb{E}_{\sigma, b, \gamma}[\mathbf{1}_{\mathcal{T}_3}|\hat{\kappa}_{J,1} - \kappa_1|^2] \lesssim N^{-2s/(2s+3)}. \quad \square \end{aligned}$$

Corollary 19. *Choosing $2^J \sim N^{1/(2s+3)}$, there exist an event $\mathcal{T}_4 = \mathcal{T}_0 \cap \mathcal{T}_3$ of high probability, i.e. $\mathbb{P}_{\sigma, b, \gamma}(\Omega \setminus \mathcal{T}_4) \lesssim N^{-2s/(2s+3)}$, such that the estimators $\hat{\mu}_J$ and $\hat{v}_{J,1}$ are uniformly bounded on \mathcal{T}_4 . Furthermore, for N big enough, we have uniformly on Θ_s and Γ*

$$\mathbb{E}_{\sigma, b, \gamma} \left[\mathbf{1}_{\mathcal{T}_4} \cdot \left(|v_1 - \hat{v}_{J,1}|^2 + \|u_1 - \hat{u}_{J,1}\|_{H^1}^2 \right) \right] \lesssim N^{-2s/(2s+3)}.$$

Proof. Note that \mathcal{T}_4 is a subset of the events from Proposition 17, Lemma 18 and the event that $\hat{\mu}_J$ is uniformly bounded from below and above (see Proposition 11). Then \mathcal{T}_4 is a high probability event and by Propositions 12 and 17, the choice $2^J \sim N^{1/(2s+3)}$ yields the claimed bound of the expectation. \square

Before we present the proof of Theorem 7 we need to another representation of the volatility estimator which allows us to bound the derivative of the estimated eigenfunction.

Lemma 20. *Set $0 < a < b < 1$. There exists a high probability event $\mathcal{T}_5 \subset \mathcal{T}_4$, $\mathbb{P}_{\sigma,b,\gamma}(\Omega \setminus \mathcal{T}_5) \lesssim N^{-2s/(2s+3)}$ such that*

$$\mathbf{1}_{\mathcal{T}_5} \cdot \hat{\sigma}_J^2(x) = \mathbf{1}_{\mathcal{T}_5} \cdot \frac{2\hat{v}_{J,1} \int_0^x \hat{u}_{J,1}(y) \hat{\mu}_J(y) dy}{(\hat{u}'_{J,1}(x) \vee c'_{a,b}) \hat{\mu}_J(x)} \wedge D$$

for a deterministic constant $c'_{a,b} > 0$ satisfying $c'_{a,b} \leq c_{a,b} \leq \inf_{x \in [a,b]} u'_1(x)$.

Proof. Recall that

$$\hat{\sigma}_J^2(x) = \frac{2\hat{v}_{J,1} \int_0^x \hat{u}_{J,1}(y) \hat{\mu}_J(y) dy}{\hat{u}'_{J,1}(x) \hat{\mu}_J(x)} \wedge D = \frac{2\hat{v}_{J,1} \int_0^x \hat{u}_{J,1}(y) \hat{\mu}_J(y) dy}{\hat{\mu}_J(x) (\hat{u}'_{J,1}(x) \vee \frac{2\hat{v}_{J,1} \int_0^x \hat{u}_{J,1}(y) \hat{\mu}_J(y) dy}{\hat{\mu}_J(x) D})}.$$

Let $m = \frac{1}{2} \inf \mu(x)$ and $M = 2 \sup \hat{\mu}_J$. By Proposition 11 $m \leq \hat{\mu}_J(x) \leq M$ for all $x \in [0, 1]$ on the event \mathcal{T}_0 . This event is especially contained in

$$\mathcal{T}_5 := \mathcal{T}_4 \cap \left\{ 4 \left\| \hat{v}_{J,1} \int_0^x \hat{u}_{J,1}(y) \hat{\mu}_J(y) dy - v_1 \int_0^x u_1(y) \mu(y) dy \right\|_\infty \leq d^2 c_{a,b} m \right\},$$

where \mathcal{T}_4 is the high probability event from Corollary 19. On \mathcal{T}_5 it holds

$$\begin{aligned} \frac{2\hat{v}_{J,1} \int_0^x \hat{u}_{J,1}(y) \hat{\mu}_J(y) dy}{D \hat{\mu}_J(x)} &\geq \frac{2v_1 \int_0^x u_1(y) \mu(y) dy - 2|\hat{v}_{J,1} \int_0^x \hat{u}_{J,1}(y) \hat{\mu}_J(y) dy - v_1 \int_0^x u_1(y) \mu(y) dy|}{D \hat{\mu}_J(x)} \\ &= \frac{\sigma^2(x) u'_1(x) \mu(x) - 2|\hat{v}_{J,1} \int_0^x \hat{u}_{J,1}(y) \hat{\mu}_J(y) dy - v_1 \int_0^x u_1(y) \mu(y) dy|}{D \hat{\mu}_J(x)} \\ &\geq \frac{d^2 c_{a,b} m}{2MD} =: c'_{a,b}. \end{aligned}$$

Furthermore, by Corollary 19, using Markov and triangle inequalities, it is easy to check that $\mathbb{P}_{\sigma,b,\gamma}(\Omega \setminus \mathcal{T}_5) \lesssim N^{-\frac{2s}{2s+3}}$, cf. estimate (25) below. \square

Proof for the volatility estimator. Set $0 < a < b < 1$. Note first that since $\mathbb{P}_{\sigma,b,\gamma}(\Omega \setminus \mathcal{T}_5) \lesssim N^{-\frac{2s}{2s+3}}$ and $\sigma, \hat{\sigma}$ are bounded we just have to verify that $\mathbb{E}_{\sigma,b,\gamma}[\mathbf{1}_{\mathcal{T}_5} \cdot \|\sigma^2 - \hat{\sigma}^2\|_{L^2([a,b])}^2] \lesssim N^{-\frac{2s}{2s+3}}$. Denote $\tilde{u}'_{J,1}(x) = \hat{u}'_{J,1}(x) \vee c'_{a,b}$ and $\tilde{\sigma}_J^2(x) = \frac{2\hat{v}_{J,1} \int_0^x \hat{u}_{J,1}(y) \hat{\mu}_J(y) dy}{\tilde{u}'_{J,1}(x) \hat{\mu}_J(x)}$. Since for $x \in [a, b]$ the functions u'_1 and μ are uniformly separated from zero, we have that on \mathcal{T}_5

$$\begin{aligned} |\sigma^2(x) - \hat{\sigma}_J^2(x)| &\leq \left| \frac{2v_1 \int_0^x u_1(y) \mu(y) dy}{u'_1(x) \mu(x)} - \frac{2\hat{v}_{J,1} \int_0^x \hat{u}_{J,1}(y) \hat{\mu}_J(y) dy}{\tilde{u}'_{J,1}(x) \hat{\mu}_J(x)} \right| \\ &= \left| \frac{2(v_1 \int_0^x u_1(y) \mu(y) dy - \hat{v}_{J,1} \int_0^x \hat{u}_{J,1}(y) \hat{\mu}_J(y) dy)}{u'_1(x) \mu(x)} - \frac{\tilde{\sigma}_J^2(x) (u'_1(x) \mu(x) - \tilde{u}'_{J,1}(x) \hat{\mu}_J(x))}{u'_1(x) \mu(x)} \right| \\ &\lesssim \left| v_1 \int_0^x u_1(y) \mu(y) dy - \hat{v}_{J,1} \int_0^x \hat{u}_{J,1}(y) \hat{\mu}_J(y) dy \right| + |\tilde{\sigma}_J^2(x)| \left| \frac{u'_1(x) \mu(x) - \tilde{u}'_{J,1}(x) \hat{\mu}_J(x)}{u'_1(x)} \right| \\ &=: A_1(x) + A_2(x). \end{aligned}$$

Observe that since $\hat{\mu}_J$ is uniformly bounded on the event \mathcal{T}_5 and since the eigenfunction \hat{u}_1 is normalized the Cauchy-Schwarz inequality grants that $\int_0^x \hat{u}_{J,1}(y) \hat{\mu}_J(y) dy$ is uniformly bounded. Hence,

$$A_1(x) = |v_1 \left(\int_0^x u_1(y) \mu(y) dy - \int_0^x \hat{u}_{J,1}(y) \hat{\mu}_J(y) dy \right) + \int_0^x \hat{u}_{J,1}(y) \hat{\mu}_J(y) dy (v_1 - \hat{v}_{J,1})|$$

$$\begin{aligned}
&\lesssim \left| \int_0^x u_1(y) \mu(y) dy - \int_0^x \widehat{u}_{J,1}(y) \widehat{\mu}_J(y) dy \right| + |v_1 - \widehat{v}_{J,1}| \\
&\lesssim \left| \int_0^x u_1(y) (\mu(y) - \widehat{\mu}_J(y)) dy \right| + \left| \int_0^x (u_1(y) - \widehat{u}_{J,1}(y)) \widehat{\mu}_J(y) dy \right| + |v_1 - \widehat{v}_{J,1}| \\
&\leq \|u_1\|_{L^2} \|\mu - \widehat{\mu}_J\|_{L^2} + \|u_1 - \widehat{u}_{J,1}\|_{L^2} \|\widehat{\mu}_J\|_{L^2} + |v_1 - \widehat{v}_{J,1}| \\
&= \|\mu - \widehat{\mu}_J\|_{L^2} + \|u_1 - \widehat{u}_{J,1}\|_{L^2} + |v_1 - \widehat{v}_{J,1}|.
\end{aligned} \tag{25}$$

Furthermore, since $\widetilde{\sigma}_J^2(x)$ is uniformly bounded on \mathcal{T}_5

$$\begin{aligned}
A_2(x) &\lesssim |\mu(x) - \widehat{\mu}_J(x)| + \frac{|\widehat{\mu}_J(x)|}{|u'_1(x)|} |u'_1(x) - \widetilde{u}'_{J,1}(x)| \\
&\lesssim |\mu(x) - \widehat{\mu}_J(x)| + |u'_1(x) - \widetilde{u}'_{J,1}(x)| \\
&\lesssim |\mu(x) - \widehat{\mu}_J(x)| + |u'_1(x) - \widetilde{u}'_{J,1}(x)|.
\end{aligned} \tag{26}$$

Consequently,

$$\begin{aligned}
\mathbb{E}_{\sigma,b,\gamma} \left[\mathbf{1}_{\mathcal{T}_5} \cdot \|\sigma^2 - \widetilde{\sigma}_J^2\|_{L^2}^2 \right] &\lesssim \mathbb{E}_{\sigma,b,\gamma} \left[\mathbf{1}_{\mathcal{T}_5} \cdot (\|A_1\|_{L^2}^2 + \|A_2\|_{L^2}^2) \right] \\
&\lesssim \mathbb{E}_{\sigma,b,\gamma} \left[\mathbf{1}_{\mathcal{T}_5} \cdot \left(\|\mu - \widehat{\mu}_J\|_{L^2}^2 + \|u_1 - \widehat{u}_{J,1}\|_{H^1}^2 + |v_1 - \widehat{v}_{J,1}|^2 \right) \right] \\
&\lesssim N^{-2s/(2s+3)}.
\end{aligned} \quad \square$$

Proof for the drift estimator. To obtain the upper bound on the drift term first note that using Bernstein's inequality we can extend the proofs of Propositions 12 and 17 to obtain

$$\mathbb{E}_{\sigma,b,\gamma} \left[\mathbf{1}_{\mathcal{T}_4} \cdot \|u_1 - \widehat{u}\|_{H^2}^2 \right] \lesssim N^{-\frac{2(s-1)}{2s+3}}. \tag{27}$$

Let $\mathcal{T}_6 = \mathcal{T}_5 \cap \{\inf_{x \in [a,b]} \widehat{u}'_{J,1}(x) \geq c_{a,b}/2\} \cap \{\|\widehat{u}_{J,1}\|_{H^2} \leq 2\|u_1\|_{H^2}\}$. By Lemma 20 and (27) we obtain that $\mathbb{P}_{\sigma,b,\gamma}(\Omega \setminus \mathcal{T}_6) \lesssim N^{-\frac{2(s-1)}{2s+3}}$. Since both b and \widehat{b} are bounded in L^2 , we can restrict the error analysis to the high probability event \mathcal{T}_6 . Recall the definition of \widetilde{b} from (12). Since $\|b\|_{L^2([a,b])} \leq D$ we have $\|\widehat{b}_J - b\|_{L^2([a,b])} \leq \|\widetilde{b}_J - b\|_{L^2([a,b])}$. Consequently, it remains to show

$$\mathbb{E}_{\sigma,b,\gamma} \left[\mathbf{1}_{\mathcal{T}_6} \cdot \|\widetilde{b}_J - b\|_{L^2([a,b])} \right] \lesssim N^{-\frac{2(s-1)}{2s+3}}.$$

On \mathcal{T}_6 , for $x \in [a, b]$ we have

$$\begin{aligned}
|\widetilde{b}_J(x) - b(x)| &\leq \left| \frac{\widehat{v}_{J,1} \widehat{u}_{J,1}(x)}{\widehat{u}'_{J,1}(x)} - \frac{\widetilde{\sigma}_J^2(x) \widehat{u}_{J,1}''(x)}{2\widehat{u}'_{J,1}(x)} - \frac{v_1 u_1(x)}{u'_1(x)} + \frac{\sigma^2(x) u_1''(x)}{2u'_1(x)} \right| \\
&\leq |u'_1(x)|^{-1} \left| \widehat{v}_{J,1} \widehat{u}_{J,1}(x) - v_1 u_1(x) + \frac{\sigma^2(x)}{2} u_1''(x) - \frac{\widetilde{\sigma}_J^2(x)}{2} \widehat{u}_{J,1}''(x) \right| \\
&\quad + \frac{|\widetilde{b}_J(x)|}{|u'_1(x)|} |u'_1(x) - \widehat{u}'_{J,1}(x)|.
\end{aligned}$$

The uniform lower bound on $|u'_1|$ yields

$$\begin{aligned}
\|\widetilde{b}_J - b\|_{L^2([a,b])}^2 &\lesssim \|\widehat{v}_{J,1} \widehat{u}_{J,1} - v_1 u_1\|_{L^2([a,b])}^2 + \|\widetilde{\sigma}_J^2 \widehat{u}_{J,1}'' - \sigma^2 u_1''\|_{L^2([a,b])}^2 \\
&\quad + \|\widetilde{b}_J\|_{L^2([a,b])}^2 \|\widehat{u}'_{J,1} - u'_1\|_{L^\infty([a,b])}^2 \\
&=: B_1 + B_2 + B_3.
\end{aligned}$$

We will estimate these three terms separately. Corollary 19 and the normalization of $\widehat{u}_{J,1}$ yield

$$\mathbb{E}_{\sigma,b,\gamma} [\mathbf{1}_{\mathcal{T}_6} B_1] \leq \mathbb{E}_{\sigma,b,\gamma} [\mathbf{1}_{\mathcal{T}} (|\widehat{v}_{J,1} - v_1|^2 \|\widehat{u}_{J,1}\|_{L^2}^2 + |v_1|^2 \|\widehat{u}_{J,1} - u_1\|_{L^2}^2)] \lesssim N^{-2s/(2s+3)}.$$

The second term can be decomposed into

$$B_2 \leq 2\|\tilde{\sigma}_J^2 - \sigma^2\|_\infty^2 \|u_1''\|_{L^2}^2 + 2\|\tilde{\sigma}_J^2\|_\infty^2 \|\hat{u}_{J,1}'' - u_1''\|_{L^2}^2.$$

From (25) and (26) we can easily verify that

$$\|\hat{\sigma}_J^2 - \sigma^2\|_\infty \lesssim |\hat{v}_{J,1} - v_1| + \|\hat{u}_{J,1} - u_1\|_{H^2} + \|\hat{\mu}_J - \mu\|_{H^1}.$$

Since $\hat{\sigma}_J^2$ is bounded by construction, we conclude

$$\mathbb{E}_{\sigma,b,\gamma}[\mathbf{1}_{\mathcal{T}_6} B_2] \leq \mathbb{E}_{\sigma,b,\gamma}[\mathbf{1}_{\mathcal{T}_6} (|\hat{v}_{J,1} - v_1|^2 + \|\hat{u}_{J,1} - u_1\|_{H^2}^2 + \|\hat{\mu}_J - \mu\|_{H^1}^2)] \lesssim N^{-2(s-1)/(2s+3)}.$$

For the last term it holds

$$\mathbb{E}_{\sigma,b,\gamma}[\mathbf{1}_{\mathcal{T}_6} B_3] \leq \mathbb{E}_{\sigma,b,\gamma}[\mathbf{1}_{\mathcal{T}_6} \|\tilde{b}_J\|_{L^2([a,b])}^2 \|\hat{u}_{J,1} - u_1\|_{H^2}^2] \lesssim N^{-2(s-1)/(2s+3)}$$

since $\|\tilde{b}_J\|_{L^2([a,b])}$ is uniformly bounded on \mathcal{T}_6 . \square

8 Proof of the lower bounds

First note that estimating the sampling distribution γ has no impact on the convergence rates, because the Laplace transform can be estimated with the parametric rate. Therefore, it suffices to use the same distribution $\gamma \in \Gamma$ for all alternatives. Throughout this section we thus fix some $\gamma \in \Gamma$ which admits a bounded Lebesgue density on $[0, T]$ for some $T > 0$.

Without loss of generality we can suppose that $(1, 0) \in \Theta_s$. To construct the alternatives, let ψ be a compactly supported wavelet in H^s with one vanishing moment. We set $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$ and denote by $K_j \subset \mathbb{Z}$ a maximal set of indices k such that $\text{supp}(\psi_{jk}) \subset [a, b]$ and $\text{supp}(\psi_{jk}) \cap \text{supp}(\psi_{jk'}) = \emptyset$ holds for all $k, k' \in K_j$, $k \neq k'$. For a constant $\delta > 0$ and all $\varepsilon = (\varepsilon_k) \in \{-1, 1\}^{|K_j|}$ we define

$$S_\varepsilon(x) = S_\varepsilon(j, x) = \left(2 + \delta \sum_{k \in K_j} \varepsilon_k \psi_{jk}(x)\right)^{-1}.$$

Choosing $\delta \sim 2^{-j(s+1/2)}$ yields $(\sqrt{2S_\varepsilon}, S'_\varepsilon) \in \Theta_s$. The corresponding diffusions $X^{(\varepsilon)}$ are defined by their generators

$$\begin{aligned} L_\varepsilon f(x) &= S_\varepsilon(x) f''(x) + S'_\varepsilon(x) f'(x), \\ \text{dom}(L_\varepsilon) &= \text{dom}(L). \end{aligned}$$

Note that for any ε the invariant measure of $X^{(\varepsilon)}$ is given by Lebesgue measure on $[0, 1]$. For $\varepsilon, \varepsilon'$ with $\|\varepsilon - \varepsilon'\|_{\ell^2} = 2$ we have

$$S_{\varepsilon'}(x) - S_\varepsilon(x) = \pm 2\delta \psi_{jk}(x) S_{\varepsilon'}(x) S_\varepsilon(x).$$

Since $S_\varepsilon, S_{\varepsilon'}$ converge uniformly to $1/2$ as $j \rightarrow \infty$, the L^2 -distances of the volatility functions and the drift functions of the alternatives ε and ε' are bounded by

$$\|2S_{\varepsilon'} - 2S_\varepsilon\|_{L^2} \gtrsim \delta, \quad \|S'_{\varepsilon'} - S'_\varepsilon\|_{L^2} \gtrsim 2^j \delta.$$

Therefore, Assouad's lemma and $\delta \sim 2^{-j(s+1/2)}$ yield for all estimators $\bar{\sigma}^2$ and \bar{b}

$$\begin{aligned} \sup_{(\sigma,b) \in \Theta_s} \mathbb{E}_{\sigma,b,\gamma} \left[\|\bar{\sigma}^2 - \sigma^2\|_{L^2([a,b])}^2 \right] &\gtrsim 2^j \delta = 2^{-2sj}, \\ \sup_{(\sigma,b) \in \Theta_s} \mathbb{E}_{\sigma,b,\gamma} \left[\|\bar{b} - b\|_{L^2([a,b])}^2 \right] &\gtrsim 2^{3j} \delta = 2^{-2(s+1)j}, \end{aligned} \tag{28}$$

provided the Kullback-Leibler divergence between the distributions of $(X_{\tau_n}^{(\varepsilon)})_{n=0,\dots,N}$ and $(X_{\tau_n}^{(\varepsilon')})_{n=0,\dots,N}$ remains uniformly bounded for all alternatives $\varepsilon, \varepsilon'$ with $\|\varepsilon - \varepsilon'\|_{\ell^2} = 2$.

To bound the Kullback-Leibler divergence, we have to take into account the random observation times. Denote the transition density of $(X_t)_{t \geq 0}$ by $p_t(x, y)dy = \mathbb{P}_{\sigma, b}(X_t = dy | X_0 = x)$ for $x, y \in [0, 1], t \geq 0$. By the independence of the observation time τ and the process X we have

$$Rf(x) = \mathbb{E}_{\sigma, b, \gamma}[f(X_\tau) | X_0 = x] = \int_0^\infty P_t f(x) \gamma(dt) = \int_0^\infty \int_0^1 p_t(x, y) f(y) dy \gamma(dt).$$

For one dimensional diffusions with bounded drift and differentiable volatility, which is uniformly separated from zero, we know that

$$p_t(x, y) \leq c_0 \left(1 + \frac{1}{\sqrt{t}}\right)$$

with $c_0 > 0$ depending only on the bounds for the drift and volatility (see Qian and Zheng [24, Thm. 1]). The assumption $\mathbb{E}[\tau^{-1/2}] < \infty$ thus ensures that

$$r(x, y) = \int_0^\infty p_t(x, y) \gamma(dt)$$

is a well defined kernel of operator R . We obtain the following generalization of Proposition 6.4 in [14]:

Lemma 21. *Assume $\mathbb{E}_\gamma[\tau^{-1/2}] < \infty$. If $(\sigma_n, b_n) \in \Theta_s$, $n \geq 0$, such that*

$$\lim_{n \rightarrow \infty} \|\sigma_n - \sigma_0\|_\infty = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \|b_n - b_0\|_\infty = 0,$$

then the corresponding kernels $r^{(n)}(x, y)dy = \mathbb{P}_{\sigma_n, b_n}(X_\tau \in dy | X_0 = x)$ satisfy

$$\lim_{n \rightarrow \infty} \|r^{(n)} - r^{(0)}\|_\infty = 0.$$

Note that the bounded Lebesgue density γ near the origin specially ensures that $\mathbb{E}_\gamma[\tau^{-1/2}] < \infty$.

Proof. Due to the bound $\|p_t^{(n)}(\cdot, \cdot)\|_\infty \lesssim 1 + t^{-1/2}$, dominated convergence yields

$$\begin{aligned} \|r^{(n)} - r^{(0)}\|_\infty &= \sup_{x, y \in [0, 1]} \left| \int_0^\infty \left(p_t^{(n)}(x, y) - p_t^{(0)}(x, y) \right) \gamma(dt) \right| \\ &\leq \int_0^\infty \|p_t^{(n)} - p_t^{(0)}\|_\infty \gamma(dt). \end{aligned}$$

By [14, Prop. 6.4] this tends to zero. \square

Exactly as in [14, Sect. 5.2], this lemma allows us to bound the Kullback-Leibler divergence by $N\|r_{\varepsilon'} - r_\varepsilon\|_{L^2([0, 1]^2)}^2$ for kernels $r_{\varepsilon'}$ and r_ε of $R_{\varepsilon'}$ and R_ε , respectively, for any $\varepsilon, \varepsilon'$ with $\|\varepsilon - \varepsilon'\|_{\ell^2} = 2$. Note that $\|r_{\varepsilon'} - r_\varepsilon\|_{L^2([0, 1]^2)}$ is the Hilbert-Schmidt norm distance $\|R - R^{\varepsilon'}\|_{HS} = \|(R^\varepsilon - R^{\varepsilon'})|_V\|_{HS}$ where

$$V = \left\{ f \in L^2([0, 1]) \mid \int_0^1 f = 0 \right\}.$$

We will bound the Hilbert-Schmidt norm by the difference of the inverses of the generators, which are, in contrast to the generators itself, bounded operators. Recall that $R = \mathcal{L}(-L)$ for the Laplace transform $\mathcal{L}(z) = \int_0^\infty e^{-tz} \gamma(dt)$, $z \geq 0$. By the functional calculus for operators the function $f(z) = \mathcal{L}(-z^{-1})$ maps $(L_\varepsilon|_V)^{-1}$ to $R^\varepsilon|_V$. Furthermore, f is uniformly Lipschitz on $(-\infty, 0)$:

Lemma 22. Suppose that $\gamma \in \Gamma$ admits a bounded Lebesgue density on $[0, T]$ for some $T > 0$. Then we have

$$c := \sup_{z < 0} \left| \frac{1}{z^2} \int_0^\infty t e^{t/z} \gamma(dt) \right| < \infty.$$

Proof. We decompose

$$\sup_{z < 0} \left| \frac{1}{z^2} \int_0^\infty t e^{t/z} \gamma(dt) \right| \leq \sup_{z < 0} \left| \frac{1}{z^2} \int_0^T t e^{t/z} \gamma(dt) \right| + \sup_{z < 0} \left| \frac{1}{z^2} \int_T^\infty t e^{t/z} \gamma(dt) \right| =: S_1 + S_2.$$

Due to the bounded Lebesgue density on $[0, T]$, we estimate the first term by substituting $s = t/z$

$$S_1 \lesssim \sup_{z < 0} z^{-2} \int_0^T t e^{t/z} dt = \sup_{z < 0} \int_{T/z}^0 s e^s ds = \int_{-\infty}^0 s e^s ds < \infty.$$

For the second term note that the function $g_a(x) = x^2 e^{-ax}$ takes maximum at $x = 2/a$ and $g(2/a) = 4a^{-2}e^{-2}$. Consequently,

$$S_2 \leq \sup_{z < 0} \int_T^\infty t g_t(|z|^{-1}) \gamma(dt) = \int_T^\infty \frac{4}{t e^2} \gamma(dt) \leq \frac{4}{T e^2} < \infty. \quad \square$$

We conclude

$$\|r_{\varepsilon'} - r_\varepsilon\|_{L^2([0,1]^2)} = \|(R^\varepsilon - R^{\varepsilon'})|_V\|_{HS} \leq c \|(L_\varepsilon|_V)^{-1} - (L_{\varepsilon'}|_V)^{-1}\|_{HS} \lesssim \delta 2^{-j} = 2^{-j(2s+3)/2},$$

by the estimate for the difference of inverses of the generators that was established in [14, Sect. 5.3]. In order to bound $N\|r_{\varepsilon'} - r_\varepsilon\|_{L^2([0,1]^2)}^2$, we thus choose j such that $2^j \sim N^{1/(2s+3)}$. In view of (28) we have proven Theorem 8. \square

9 Proof for the adaptive estimator

In order to show that Lepski's method works, we need the following concentration result. It slightly generalizes the corresponding concentration inequalities by Nickl and Söhl [23, Theorems 10 and 11] for a low-frequently observed reflected diffusion to random sampling times.

Proposition 23. Grant Assumptions 1 and 3 with $s > 5/2$ and $\gamma \in \Gamma$, $\mathbb{E}_\gamma[\tau^{-1/2}] \leq D$. There is a constant $c > 0$ depending only on d, D, I and α , such that, for any $\kappa > 0, N \in \mathbb{N}$ and any $f \in L^2(\mathbb{R}) \cap L^\infty(\mathbb{R}), g \in L^2(\mathbb{R}^2) \cap L^\infty(\mathbb{R}^2)$:

$$\mathbb{P}_{\sigma, b, \gamma} \left(\left| \sum_{n=0}^N (f(X_{\tau_n}) - \mathbb{E}_{\sigma, b, \gamma}[f(X_0)]) \right| > \kappa \right) \lesssim \exp \left(-c \min \left\{ \frac{\kappa^2}{N \|f\|_{L^2}^2}, \frac{\kappa}{(\log N) \|f\|_\infty} \right\} \right)$$

and

$$\begin{aligned} \mathbb{P}_{\sigma, b, \gamma} \left(\left| \sum_{n=0}^{N-1} (g(X_{\tau_n}, X_{\tau_{n+1}}) - \mathbb{E}_{\sigma, b, \gamma}[g(X_0, X_{\tau_1})]) \right| > \kappa \right) \\ \lesssim \exp \left(-c \min \left\{ \frac{\kappa^2}{N \|g\|_{L^2}^2}, \frac{\kappa}{(\log N) \|g\|_\infty} \right\} \right). \end{aligned}$$

Proof. The conditions of the Markov chain concentration result by Adamczak [1, Theorem 6] have to be verified. This can be done along the lines of the proofs in [23] using Lemma 4 and noting that the transition density of the time-changed chain $(X_{\tau_n})_{n \geq 1}$ is given by $p_\gamma(x, y) = \int_0^\infty p_t(x, y) \gamma(dt)$ where $p_t(x, y)$ denotes the transition density of the diffusion $(X_t)_{t \geq 0}$. The condition $s > 5/2$ ensures that the transition density p_γ is bounded from below uniformly on $[0, 1]^2$. Indeed, $p_\gamma(x, y) \geq K\gamma(I) \geq K\alpha$, where K is the uniform lower bound on $\inf_{t \in I} p_t$ obtained in [23, Proposition 9]. Since $\|p_t\|_\infty \lesssim 1 + t^{-1/2}$, the condition $\mathbb{E}_\gamma[\tau^{-1/2}] < \infty$ ensures a uniform upper bound on p_γ . \square

To analyze the performance of $\tilde{\sigma}^2$, we first decompose its estimation error into a deterministic and a stochastic error term. In what follows, $C = C(d, D, I, \alpha)$ denotes a numeric constant which may vary from line to line. We deduce from the proof of Theorem 7 on the there defined event \mathcal{T}_5 , that for any $J \in \mathcal{J}_N$

$$\begin{aligned} \|\hat{\sigma}_J^2 - \sigma^2\|_{L^2} &\leq C(\|\mu - \hat{\mu}_J\|_{L^2} + \|u_1 - \hat{u}_{J,1}\|_{H^1} + |v_1 - \hat{v}_{J,1}|) \\ &\leq C(\|\mu - \hat{\mu}_J\|_{L^2} + \|u_1 - \hat{u}_{J,1}\|_{H^1} + |\kappa_1 - \hat{\kappa}_{J,1}| + |\mathcal{L}_\gamma(-v_1) - \hat{\mathcal{L}}_\gamma(-v_1)|) \\ &\leq D_J + S_J, \end{aligned} \tag{29}$$

where

$$\begin{aligned} D_J &:= C(\|(I - \pi_J)\mu\|_{L^2} + \|u_1 - u_{J,1}\|_{H^1} + |\kappa_1 - \kappa_{J,1}|), \\ S_J &:= C(\|\pi_J\mu - \hat{\mu}_J\|_{L^2} + \|u_{J,1} - \hat{u}_{J,1}\|_{H^1} + |\kappa_{J,1} - \hat{\kappa}_{J,1}| + |\mathcal{L}_\gamma(-v_1) - \hat{\mathcal{L}}_\gamma(-v_1)|). \end{aligned}$$

Due to the smoothness of the invariant measure, Jackson's inequality and Proposition 12, there is some $\beta > 0$, depending on ψ, d and D such that

$$D_J \leq \beta 2^{-Js}.$$

We need that S_J concentrates around zero. Recalling the definition of the residual vector

$$r = (\hat{R}_J - R_J)u_{J,1} + \kappa_{J,1}(G_J - \hat{G}_J)u_{J,1},$$

Bernstein's inequality and Theorem 26 on generalized symmetric eigenvalue problems yield, on the event \mathcal{T}_2 from Proposition 17, that

$$\|u_{J,1} - \hat{u}_{J,1}\|_{H^1} + |\kappa_{J,1} - \hat{\kappa}_{J,1}| \leq C2^J \|u_{J,1} - \hat{u}_{J,1}\|_{L^2} + |\kappa_{J,1} - \hat{\kappa}_{J,1}| \leq \|r\|_{L^2} (C2^J + 1).$$

Corollary 24. *Under the conditions of Proposition 23, for any $\tau > 1$ there exist $\eta_1, \eta_2, \eta_3 > 1$, such that, for all J with $2^J \lesssim \frac{N}{(\log N)^2 \log \log N}$, we have*

$$\mathbb{P}_{\sigma,b,\gamma} \left(\|\pi_J\mu - \hat{\mu}_J\|_{L^2} > 2^{\frac{J}{2}} \eta_1 \sqrt{\frac{\log \log N}{N}} \right) \lesssim (\log N)^{-\tau}, \tag{30}$$

$$\mathbb{P}_{\sigma,b,\gamma} \left(\|r\|_{L^2} > 2^{\frac{J}{2}} \eta_2 \sqrt{\frac{\log \log N}{N}} \right) \lesssim (\log N)^{-\tau}, \tag{31}$$

$$\mathbb{P}_{\sigma,b,\gamma} \left(|\mathcal{L}_\gamma(-v_1) - \hat{\mathcal{L}}_\gamma(-v_1)| > \eta_3 \sqrt{\frac{\log \log N}{N}} \right) \lesssim (\log N)^{-\tau}. \tag{32}$$

In particular, there is a $\Lambda > 0$ such that $\mathbb{P}_{\sigma,b,\gamma}(4S_J > s_J) \lesssim (\log N)^{-\tau}$ for $s_J = s_J(\Lambda)$ from (14).

Proof. Fix $\tau > 1$. Since $\|\psi_\lambda\|_\infty \lesssim 2^{|\lambda|/2}$, for $|\lambda| \leq J$, using Proposition 23 we obtain

$$\begin{aligned} \mathbb{P}_{\sigma,b,\gamma} \left(|\langle \psi_\lambda, \mu - \mu_N \rangle| > \eta_1 \sqrt{\frac{\log \log N}{N}} \right) &\lesssim \exp \left(-c \min \left\{ \frac{\eta_1^2 N (\log \log N)}{N \|\psi_\lambda\|_{L^2}^2}, \frac{\eta_1 \sqrt{N (\log \log N)}}{(\log N) \|\psi_\lambda\|_\infty} \right\} \right) \\ &\lesssim \exp \left(-c \eta_1 \min \left\{ \log \log N, \frac{\sqrt{N (\log \log N)}}{(\log N) 2^{J/2}} \right\} \right) \\ &\lesssim (\log N)^{-c\eta_1} \lesssim (\log N)^{-\tau}, \end{aligned}$$

for some η_1 big enough. Applying a usual chaining argument, this concentration inequality carries over to $\max_{|\lambda| \leq J} |\langle \psi_\lambda, \mu - \mu_N \rangle|$, cf. [7, Theorem 2.1] and [23, Theorem 12]. Since $\|\mu_J - \hat{\mu}_J\|_{L^2}^2 = \sum_{|\lambda| \leq J} |\langle \psi_\lambda, \mu - \mu_N \rangle|^2$, it follows that

$$\mathbb{P}_{\sigma,b,\gamma} \left(\|\pi_J\mu - \hat{\mu}_J\|_{L^2}^2 > \eta_1^2 2^J \frac{\log \log N}{N} \right) \lesssim \mathbb{P}_{\sigma,b,\gamma} \left(\max_{|\lambda| \leq J} |\langle \psi_\lambda, \mu - \mu_N \rangle|^2 > \eta_1^2 \frac{\log \log N}{N} \right) \lesssim (\log N)^{-\tau}.$$

To prove (31), note first that since $|\kappa_{J,1}| \leq 1$, we have

$$\|r\|_{L^2} \leq \|(\widehat{R}_J - R_J)u_{J,1}\|_{L^2} + \|(G_J - \widehat{G}_J)u_{J,1}\|_{L^2}.$$

By Proposition 12 $\|u_{J,1}\|_{L^2}, \|u_{J,1}\|_\infty \lesssim 1$ holds for J big enough. Using the second inequality in Proposition 23, we obtain

$$\begin{aligned} \mathbb{P}_{\sigma,b,\gamma} \left(|\langle \psi_\lambda, (\widehat{R}_J - R_J)u_{J,1} \rangle| > \eta_2 \sqrt{\frac{\log \log N}{N}} \right) \\ \lesssim \exp \left(-c\eta_2 \min \left\{ \frac{N(\log \log N)}{N}, \frac{\sqrt{N(\log \log N)}}{(\log N)^{2J/2}} \right\} \right) \lesssim (\log N)^{-C\eta_2} \lesssim (\log N)^{-\tau}, \end{aligned}$$

for η_2 big enough. Since $\|(\widehat{R}_J - R_J)u_{J,1}\|_{L^2}^2 = \sum_{|\lambda| \leq J} |\langle \psi_\lambda, (\widehat{R}_J - R_J)u_{J,1} \rangle|^2$, we conclude again that

$$\mathbb{P}_{\sigma,b,\gamma} \left(\|(\widehat{R}_J - R_J)u_{J,1}\|_{L^2} > \eta_2 2^{\frac{J}{2}} \sqrt{\frac{\log \log N}{N}} \right) \lesssim (\log N)^{-\tau}.$$

Arguing similarly we deduce also $\mathbb{P}_{\sigma,b,\gamma} \left(\|(G_J - \widehat{G}_J)u_{J,1}\|_{L^2} > \eta_2 2^{\frac{J}{2}} \sqrt{\frac{\log \log N}{N}} \right) \lesssim (\log N)^{-\tau}$ and thus (31) holds.

The concentration inequality (32) follows from the classical Bernstein inequality. Indeed, we have

$$\widehat{\mathcal{L}}_\gamma(-v_1) - \mathcal{L}_\gamma(-v_1) = \frac{1}{N} \sum_{n=1}^N \xi_n \quad \text{with} \quad \xi_n := e^{v_1 \Delta_n} - \mathbb{E}_\gamma[e^{v_1 \Delta_n}],$$

where, by Assumption 1 the random variables ξ_n are independent, centered and deterministically bounded by 2 (because $v_1 < 0$). Since $\text{Var}_\gamma(\xi_n) \leq \mathcal{L}_\gamma(-2v_1) \leq 1$, we can choose η_3 uniformly for all $\gamma \in \Gamma$. \square

We can now prove the convergence rate for the adaptive estimator.

Proof of Theorem 9. Let us introduce the oracle projection level

$$J^* := \min \{J \in \mathcal{J}_N : \beta 2^{-Js} < s_J/4\}.$$

By the choice of \mathcal{J}_N we deduce $2^{J^*} \sim (N/\log \log N)^{1/(2s+3)}$ and $s_{J^*}^2 \sim (\log \log N/N)^{2s/(2s+3)}$. Since the number of elements in \mathcal{J}_N is of order $\log N$, Proposition 23 yields $\mathbb{P}_{\sigma,b,\gamma}(\mathcal{A}_N) \rightarrow 1$ for the event

$$\mathcal{A}_N := \{\forall J \in \mathcal{J}_N : 4S_J \leq s_J\} \cap \mathcal{T}_6$$

with \mathcal{T}_6 from the proof of Theorem 7. Due to the decomposition (29), on \mathcal{A}_N we have for every $J \in \mathcal{J}_N$:

$$\|\widehat{\sigma}_J^2 - \sigma^2\|_{L^2} \leq D_J + S_J \leq \beta 2^{-Js} + s_J.$$

Hence, for all $J \geq J^*$, $J \in \mathcal{J}_N$, we obtain

$$\|\widehat{\sigma}_J^2 - \sigma^2\|_{L^2[a,b]} \leq \frac{1}{2}s_J,$$

and thus, by the triangle inequality,

$$\|\widehat{\sigma}_J^2 - \widehat{\sigma}_{J^*}^2\|_{L^2[a,b]} \leq s_J,$$

for all $J \geq J^*$, $J \in \mathcal{J}_N$. By definition of \widehat{J} , we conclude that $\widehat{J} \leq J^*$ on the event \mathcal{A}_N . We conclude that

$$\|\widetilde{\sigma}^2 - \sigma^2\|_{L^2[a,b]} \leq \|\widehat{\sigma}_{\widehat{J}}^2 - \widehat{\sigma}_{J^*}^2\|_{L^2[a,b]} + \|\widehat{\sigma}_{J^*}^2 - \sigma^2\|_{L^2[a,b]} \leq s_{J^*} + \frac{1}{2}s_{J^*} \leq \frac{3}{2}s_{J^*}. \quad \square$$

A Stability of the eigenvalue problems

A.1 Compact, self-adjoint, positive-definite operators

Theorem 25. Consider T a compact, self-adjoint and positive-definite operator on some Hilbert space $\mathcal{H} = (H, \|\cdot\|)$. Denote its eigenpairs by $(\lambda_i, x_i)_{i=1,2,\dots}$, normalized so that $\|x_i\| = 1$ and ordered decreasingly with respect to the eigenvalues. Let $V \subset H$ be a finite dimensional subspace of H , and π the orthogonal projection on V . Assume that the biggest eigenvalue λ_1 is simple and that

$$\|(I - \pi)x_1\| < \frac{\lambda_1 - \lambda_2}{6\lambda_1}.$$

Consider the projected operator $\pi T \pi$ and denote its normalized, ordered decreasingly, eigenpairs by $(\lambda_i^V, x_i^V)_{i=1,2,\dots,\dim(V)}$. Then

$$|\lambda_1 - \lambda_1^V| + \|x_1 - x_1^V\| \leq C \|(I - \pi)x_1\|$$

holds, where the constant C depends only on the size of the spectral gap $\lambda_1 - \lambda_2$ and the first eigenvalue λ_1 .

Proof. Since T is self-adjoint and positive-definite $\|T\| = \sup_{x \in H} \frac{\langle Tx, x \rangle}{\|x\|^2} = \lambda_1$. By the variational characterization of the eigenvalues

$$\lambda_i^V = \sup_{\substack{S \subset V \\ \dim(S)=i}} \inf_{y \in S} \frac{\langle y, Ty \rangle}{\|y\|^2} \leq \sup_{\substack{S \subset H \\ \dim(S)=i}} \inf_{y \in S} \frac{\langle y, Ty \rangle}{\|y\|^2} = \lambda_i. \quad (33)$$

Furthermore

$$\begin{aligned} \lambda_1 - \lambda_1^V &\leq \frac{\langle (\lambda_1 - \pi T \pi)(\pi x_1), \pi x_1 \rangle}{\|\pi x_1\|^2} = \frac{\langle \pi T (I - \pi)x_1, \pi x_1 \rangle}{\|\pi x_1\|^2} \\ &\leq \frac{\|\pi T (I - \pi)x_1\|}{\|\pi x_1\|} \leq \|T\| \frac{\|(I - \pi)x_1\|}{\|\pi x_1\|} \\ &\leq \|T\| \frac{\|(I - \pi)x_1\|}{1 - \|(I - \pi)x_1\|}. \end{aligned}$$

Since $|\lambda_1 - \lambda_1^V| \leq 2\|T\|$, from the inequality $\frac{z}{1-z} \wedge 2 \leq 3z$ for $z = \|(I - \pi)x_1\|$ follows that

$$|\lambda_1 - \lambda_1^V| \leq 3\|T\| \|(I - \pi)x_1\|.$$

Since by (33) holds $\lambda_2^V \leq \lambda_2$ and $\|T\| \|(I - \pi)x_1\| < \frac{\lambda_1 - \lambda_2}{6}$ we have

$$\begin{aligned} |\lambda_1^V - \lambda_2^V| &\geq \lambda_1^V - \lambda_2 = |\lambda_1 - \lambda_2| - |\lambda_1 - \lambda_1^V| \\ &\geq \lambda_1 - \lambda_2 - 3\|T\| \|(I - \pi)x_1\| \geq \frac{1}{2}(\lambda_1 - \lambda_2). \end{aligned}$$

Consequently the projected operator $\pi T \pi$ has a spectral gap of size $\rho \geq \frac{\lambda_1 - \lambda_2}{2}$ and in particular the eigenvalue λ_1^V is simple. Define the residual vector $r = (\pi T \pi - T)x_1$. Then

$$\begin{aligned} \|r\| = \|(\pi T \pi - T)x_1\| &\leq \|\pi T \pi x_1 - \pi T x_1\| + \lambda_1 \|\pi x_1 - x_1\| \\ &\leq (\|T\| + \lambda_1) \|(I - \pi)x_1\|. \end{aligned}$$

Consequently, in order to prove $\|x_1 - x_1^V\| \leq C \|(I - \pi)x_1\|$, it suffices to justify that

$$\|x_1 - x_1^V\| \leq \frac{3\rho^2}{2\sqrt{2}} \|r\|$$

Let P be the spectral projection on the eigenspace of operator $\pi T \pi$ corresponding to the eigenvalue λ_1^V . Let $R(\pi T \pi, z) = (\pi T \pi - z)^{-1}$ be the resolvent operator. Using Cauchy's integral representation of the spectral projection (see Lemma 6.4 from [9]) and $|\lambda_1 - \lambda_1^V| \leq \rho$ we find

$$\begin{aligned} \|x_1 - Px_1\| &= \frac{1}{2\pi} \left\| \oint_{S(\lambda_1, 3\rho/2)} \frac{R(\pi T \pi, z)}{\lambda_1 - z} dz (\pi T \pi - T) x_1 \right\| \\ &\leq \frac{3\rho}{2} \|r\| \sup_{z \in S(\lambda_1, 3\rho/2)} \|R(\pi T \pi, z)\|. \end{aligned}$$

Since operator $\pi T \pi$ is self adjoint on \mathcal{H} we know that (see Proposition 2.32 from [9]) $\|R(\pi T \pi, z)\| = (\text{dist}(z, \sigma(\pi T \pi)))^{-1}$. Consequently

$$\sup_{z \in S(\lambda_1, 3\rho/2)} \|R(\pi T \pi, z)\| = \sup_{z \in S(\lambda_1, 3\rho/2)} (\text{dist}(z, \sigma(\pi T \pi)))^{-1} \leq \frac{\rho}{2}.$$

It remains to bound the distance between the eigenvectors. Since x_1 and x_1^V are normalized

$$\begin{aligned} \|x_1^V - x_1\|^2 &= 2 - 2\langle x_1^V, x_1 \rangle \leq 2 - 2\langle x_1^V, x_1 \rangle^2 \\ &= 2(1 + \langle x_1^V, x_1 \rangle)(1 - \langle x_1^V, x_1 \rangle) = 2\|x_1 - \langle x_1^V, x_1 \rangle x_1^V\|^2. \end{aligned}$$

Since λ_1^V is simple, the right hand side is equal to $2\|x_1 - Px_1\|^2$. \square

A.2 Generalized symmetric eigenvalue problems.

In this section we want to sketch the a posteriori technique of solving generalized symmetric eigenvalue problems (GSEP). GSEPs have been studied extensively in chapter VI of [30]. For the error analysis in the case of standard matrix eigenvalue problems we refer to Chapter 1 of [9] or Chapter V of [30]. A particularly useful reference for various eigenvalue problems is [5].

Consider $A, B \in \mathbb{R}^{n \times n}$ real, symmetric matrices with B positive definite. We call a pair $(\lambda, x) \in \mathbb{R} \times (\mathbb{R}^n \setminus \{0\})$ an eigenpair of the generalized symmetric eigenvalue problem (GSEP) for matrices A, B if

$$Ax = \lambda Bx. \quad (34)$$

Furthermore we adapt the notation of the standard eigenvalue problems calling λ the eigenvalue and x the eigenvector. An eigenpair is normalized if $\|x\| = 1$, where $\|x\| = (\sum_{i=1}^n x_i^2)^{\frac{1}{2}}$ is the Euclidean norm on \mathbb{R}^n .

Using Cholesky decomposition of matrix $B = DD^*$ one can reduce the generalized problem (34) to the standard eigenvalue problem for matrix $D^{-1}AD^{-*}$. We deduce that problem (34) has n solutions $(\lambda_i, x_i)_{i=1, \dots, n}$, all eigenvalues are real and we can ordered the eigenpairs with respect to the eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. Furthermore corresponding eigenvectors $(x_i)_{i=1, \dots, n}$ form a B -orthogonal basis of \mathbb{R}^n .

Consider now perturbed matrices \tilde{A}, \tilde{B} with \tilde{B} positive definite and the corresponding GSEP:

$$\tilde{A}\tilde{x} = \tilde{\lambda}\tilde{B}\tilde{x}. \quad (35)$$

We want to formulate error bounds between $(\tilde{\lambda}_1, \tilde{x}_1)$ and (λ_1, x_1) . To that purpose form the residual vector

$$r = A\tilde{x}_1 - \tilde{\lambda}_1 B\tilde{x}_1 = (A - \tilde{A})\tilde{x}_1 + \tilde{\lambda}_1(\tilde{B} - B)\tilde{x}_1.$$

The standard a posteriori procedure is to find a matrix $E = E(\tilde{\lambda}_1, \tilde{x}_1)$ such that

$$\begin{aligned} (A + E)\tilde{x}_1 &= \tilde{\lambda}_1 B\tilde{x}_1, \\ \|E\| &= \|r\|. \end{aligned} \quad (36)$$

Since we replaced in (36) the perturbed matrix \tilde{B} by B , the final step is to reduce (36) and (34) to the standard eigenvalue problems using the Cholesky decomposition of B . Then we can apply the standard error bounds expressed in terms of the perturbation matrix E . We obtain

Theorem 26. *There exists a normalized eigenpair (λ_i, x_i) , $1 \leq i \leq n$ such that*

$$\begin{aligned} |\lambda_i - \tilde{\lambda}_1| &\leq \|B^{-1}\| \|r\|, \\ \|x_i - \tilde{x}_1\| &\leq \frac{2\sqrt{2\kappa(B)}}{\delta(\lambda_i)} \|B^{-1}\| \|r\|. \end{aligned}$$

where $\kappa(B) = \|B\| \|B^{-1}\|$ is the condition number of matrix B and $\delta(\lambda_i)$ is the so called localizing distance, i.e. $\delta(\lambda_i) = \min_{j \neq i} |\lambda_j - \tilde{\lambda}_1|$.

The disadvantage of the above procedure is that we obtain an existence result that gives no information how the eigenpair (λ_i, x_i) is related to (λ_1, x_1) . This is a typical downside for a posteriori methods that are supposed to provide information how far the calculated solution is from the nearest exact solution but are not intended to compare ordered eigenpairs. A helpful result is the absolute Weyl theorem for generalized hermitian definite matrix pairs, established by Y. Nakatsukasa [21]. For readers convenience we state below the theorem in the form presented in [22, Theorem 8.3].

Theorem 27. *Let $\lambda_1 \geq \dots \geq \lambda_n$ and $\tilde{\lambda}_1 \geq \dots \geq \tilde{\lambda}_n$ be respectively exact and approximated eigenvalues of problems (34) and (35). Denote $\Delta A = A - \tilde{A}$ and $\Delta B = B - \tilde{B}$. Then*

$$\begin{aligned} |\lambda_i - \tilde{\lambda}_i| &\leq \|\tilde{B}^{-1}\| \|\Delta A - \lambda_i \Delta B\|, \\ |\lambda_i - \tilde{\lambda}_i| &\leq \|B^{-1}\| \|\Delta A - \tilde{\lambda}_i \Delta B\|, \end{aligned}$$

for all $i = 1, \dots, n$.

References

- [1] Adamczak, R. (2007). A tail inequality for suprema of unbounded empirical processes with applications to markov chains. *Electron. J. Probab.*, 13:1000–1034.
- [2] Ait-Sahalia, Y. (2010). *Econometrics of Diffusion Models*. John Wiley & Sons, Ltd.
- [3] Ait-Sahalia, Y. and Mykland, P. A. (2003). The effects of random and discrete sampling when estimating continuous-time diffusions. *Econometrica*, 71(2):483–549.
- [4] Ait-Sahalia, Y. and Mykland, P. A. (2004). Estimators of diffusions with randomly spaced discrete observations: a general theory. *Ann. Statist.*, 32(5):2186–2222.
- [5] Bai, Z. and Li, R. (2000). Stability and accuracy assessments. *Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, editors, Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*.
- [6] Ball, C. A. and Roma, A. (1998). Detecting mean reversion within reflecting barriers: application to the European Exchange Rate Mechanism. *Applied Mathematical Finance*, 5:1–15.
- [7] Baraud, Y. (2010). A bernstein-type inequality for suprema of random processes with applications to model selection in non-gaussian regression. *Bernoulli*, 16(4):1064–1085.
- [8] Bass, R. (1995). *Probabilistic Techniques in Analysis*. Springer-Verlag, New York.
- [9] Chatelin, F. (1983). *Spectral approximation of linear operators*. Academic Press, New York.
- [10] Chen, X., Hansen, L. P., and Scheinkman, J. (2009). Nonlinear principal components and long-run implications of multivariate diffusions. *The Annals of Statistics*, 37:4279–4312.

- [11] Ciesielski, Z. (1963). Properties of the orthonormal Franklin system. *Studia Mathematica*, XXIII:141–157.
- [12] Ciesielski, Z. and Figiel, T. (1982). Spline approximation and Besov spaces on compact manifolds. *Studia Mathematica*, LXXV:13–36.
- [13] Fan, J. (2005). A selective overview of nonparametric methods in financial econometrics. *Statist. Sci.*, 20(4):317–357. With comments and a rejoinder by the author.
- [14] Gobet, E., Hoffmann, M., and Reiß, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *The Annals of Statistics*, 32.
- [15] Hansen, L. P., Scheinkman, J. A., and Touzi, N. (1998). Spectral methods for identifying scalar diffusions. *Journal of Econometrics*, 86:1–32.
- [16] Iglehart, D. L. and Whitt, W. (1970). Multiple Channel Queues in Heavy Traffic. *Advances in Applied Probability*, 2:150–177.
- [17] Karlin, S. and Taylor, H. (1981). *A second course in stochastic processes*. Academic Press, London and New York.
- [18] Kingman, J. F. C. (1962). On Queues in Heavy Traffic. *Journal of the Royal Statistical Society*, 24:383–392.
- [19] Krugman, P. R. (1991). Target zones and exchange rate dynamics. *The Quarterly Journal of Economics*, 106:669–682.
- [20] Lepskiĭ, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatnost. i Primenen.*, 35(3):459–470.
- [21] Nakatsukasa, Y. (2010). Absolute and relative Weyl theorems for generalized eigenvalue problem. *Linear Algebra Appl.*, 432:242–248.
- [22] Nakatsukasa, Y. (2011). *Algorithms and Perturbation Theory for Matrix Eigenvalue Problems and the Singular Value Decomposition*. PhD thesis, Davis, CA, USA. AAI3482268.
- [23] Nickl, R. and Söhl, J. (2015). Nonparametric bayesian posterior contraction rates for discretely observed scalar diffusions. *arXiv preprint arXiv:1510.05526*.
- [24] Qian, Z. and Zheng, W. (2002). Sharp bounds for transition probability densities of a class of diffusions. *C. R. Acad. Sci. Paris*, I 335:953–957.
- [25] Ricciardi, L. M. (1986). Stochastic Population Theory: Diffusion Processes. *Mathematical Ecology Biomathematics*, 17:191–238.
- [26] Rozkosz, A. and Słomiński, L. (1997). On stability and existence of solutions of SDEs with reflection at the boundary. *Stochastic processes and their applications*, 68:285–302.
- [27] Shepp, L. A. and Shiryaev, A. N. (1993). A new look at pricing of the "Russian Option". *Theory Probab. Appl.*, 39, No. 1:103–119.
- [28] Söhl, J. and Trabs, M. (2014). Adaptive confidence bands for markov chains and diffusions: Estimating the invariant measure and the drift. *arXiv preprint arXiv:1412.7103*.
- [29] Spokoiny, V. G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.*, 24(6):2477–2498.
- [30] Stewart, G. and Sun, J. (1990). *Matrix Perturbation Theory*. Academic Press, Inc.
- [31] Svensson, L. E. O. (1990). The term structure of interest rate differentials in a target zone: Theory and Swedish data. *National Bureau of Economic Research*, No. 3374.

- [32] Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.
- [33] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics.